ATTORNEY'S DOCKET NUMBER: 2007651-0001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Jack *et al.*	**Examiner:**	Hutson, Richard G.
Serial No.:	10/089,027	**Art Unit:**	1652
Filing Date:	March 26, 2002	**Confirmation No.:**	9409
Title:	INCORPORATION OF MODIFIED NUCLEOTIDES BY ARCHAEON ANA POLYMERASES AND RELATED METHODS		

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

ELECTRONICALLY FILED VIA EFS WEB

Sir:

AMENDED APPEAL BRIEF UNDER 37 C.F.R. § 41.37

Appellants hereby appeal to the Board of Patent Appeals and Interferences (the "Board") from the Examiner's final rejection of pending claims 32-42 of the above-referenced application.

A final Office Action was mailed on April 21, 2009. A Notice of Appeal was filed on July 21, 2009. An original Appeal Brief was filed on December 23, 2009 with a Petition under 37 C.F.R. § 1.136 for four month extension of time. An electronic payment of the $270.00 fee for filing an appeal brief under 37 C.F.R 41.20(b)(2) and $865 fee for an extension of time was filed concurrently with the original Appeal Brief.

A Notice of Non-Compliant Appeal Brief was mailed on January 25, 2010, setting a deadline of February 25, 2010 for filing an Amended Appeal Brief that complies with the requirements of 37 C.F.R. § 41.37. Thus, the present Amended Appeal Brief is timely filed on February 2, 2010.

Applicant believes that no further petitions and fees are required for this Appeal Brief to be entered. Please consider this a conditional petition for any additional extensions, if needed, and please charge any additional fees or credit any overpayments that may be required to our Deposit Account No. 03-1721 referencing attorney docket number 2007651-0001.

REAL PARTY IN INTEREST

As a result of an assignment by the inventors in the present application, the real party in interest in this application is New England Biolabs, Inc. The assignment to New England Biolabs, Inc. was recorded in the Patent and Trademark Office at Reel 012921, Frame 0103.

RELATED APPEALS AND INTERFERENCES

No other appeals or interferences are known to Appellants, Appellants' legal representative, or Appellants' assignee that will directly affect or be directly affected by the Board's decision in this appeal. Similarly, no such appeals or interferences are known that may have a bearing on the Board's decision in this appeal.

STATUS OF CLAIMS

The application was filed with 31 claims. Various claims were amended and/or cancelled in Amendments filed on June 7, 2002, April 18, 2005 (not entered), August 15, 2005, May 4, 2006 (not entered), May 30, 2006 (not entered), July 3, 2006, and February 24, 2007. Pending claims 2-4, 13-22, and 27-31 were canceled in the amendment filed October 31, 2007, and new claims 32-43 were presented. Claim 32 was amended in an Amendment filed on August 29, 2008 (not entered) and an Amendment filed on January 9, 2009. Claims 32-42 were finally rejected in an Office Action mailed April 21, 2009. Claim 43 is objected to for depending on rejected claims 32 and 33.

Thus, claims 1-31 are canceled, claims 32-42 are rejected, and claim 43 is objected to. The rejection of claims 32-42 is hereby appealed. A listing of the claims is provided in the attached **Claims Appendix**.

STATUS OF AMENDMENTS

There are no outstanding amendments to the claims.

SUMMARY OF CLAIMED SUBJECT MATTER

DNA polymerases are enzymes that catalyze polymerization of nucleotides into a DNA strand. The present invention encompasses the finding that a certain class of DNA polymerases has the ability to incorporate a particular modified type of nucleotide, acyclonucleotides, into DNA strands. The present claims therefore recite use of DNA polymerases from that class (defined by the present of a particular amino acid motif whose presence is shown to correlate with the activity) to incorporate acyclonucleotides into a polynucleotide chain.

Independent claim 32 and dependent claims 33-43 specifically recite methods comprising steps of providing a DNA polymerase of the relevant class, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide, and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides. The claims require that the utilized DNA polymerase be a member of the relevant class of DNA polymerases by specifying both a level of overall sequence identity to a member of the class and the presence of the correlated motif. Specifically, the claims specify that the DNA polymerase as an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to three (i.e., 0-3) amino acid substitutions as compared with the SEQ ID NO.

The claimed methods are described, inter alia, in original claims 9, 10; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3; and Table 3 on pages 20-21 of the specification. Support for claim 32 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3 and lines 10-16; and Table 3 on pages 20-21. Support for claim 33 is found in the specification as originally filed, inter alia, in original claim 10 and at page 19, lines 18-20. Support for claim 34 found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 35 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 36 is found in the specification as originally filed, inter alia, in

original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 37 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 38 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 39 is found in the specification as originally filed, inter alia, in original claim 19 Support for claim 40 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 41 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 42 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 43 is found in the specification as originally filed, inter alia, in original claims 13 and 18.

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The grounds of rejection to be reviewed on appeal are:

(1) Are claims 32-42 invalid for lack of written description under 35 U.S.C. § 112?

(2) Are claims 32 -42 invalid for lack of enablement under 35 U.S.C. § 112?

ARGUMENT

Claims 32 and 39 stand or fall together and each of claims 33, 34, 35, 36, 37, 38, 40, 41, and 42 stands or falls alone for grounds of rejection (1) and (2) to be reviewed upon appeal, as indicated below.

Ground of Rejection 1:

Claims 32 and 39 are not invalid for lack of written description

Pending claims 32-42 stand rejected for lack of written description. The Examiner states that claims 32-42 contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention. This rejection is respectfully traversed. Reconsideration and withdrawal is requested.

The written description requirement serves both to satisfy the inventor's obligation to disclose the technologic knowledge upon which the patent is based, and to demonstrate that the patentee was in possession of the invention that is claimed. *Capon v. Eshhar*, 418 F.3d 1349, 1357 (Fed. Cir. 2005). To satisfy the written description requirement, the applicant does not have to utilize any particular form of disclosure to describe the subject matter claimed, but the description must clearly allow persons of ordinary skill in the art to recognize that he or she invented what is claimed. *Carnegie Mellon Univ. v. Hoffmann La Roche Inc.*, 541 F.3d 1115, 1122 (Fed. Cir. 2008) (quoting *In re Alton*, 76 F.3d 1168 (Fed. Cir. 1996)). In other words, the applicant must 'convey with reasonably clarity to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention,' and demonstrate that by disclosure in the specification of the patent. *Id.* Such disclosure need not recite the claimed invention *in haec verba*, but it must do more than merely disclose that which would render the claimed invention obvious. *Univ. of Rochester v. G.D. Searle & Co.*, 358 F.3d 916, 923 (Fed. Cir. 2004). The descriptive text needed to meet the written description requirement varies with the nature and scope of the invention at issue, and with the scientific and technologic knowledge already in existence. *Capon*, 418 F.3d at 1357.

The present claims recite methods comprising steps of providing a particular type of DNA polymerase, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides. Claim 32 specifies that the DNA polymerase has an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. The recited 15 amino acid motifs are shown in Table 3 of the specification at pages 20-21.

The Examiner has maintained that the written description requirement is not met for the scope of DNA polymerases encompassed by the claims. Appellants explain below that a structure/function relationship has been established for the DNA polymerases recited in the claimed methods, and that written description for the claims is more than satisfied under *Invitrogen Corp. v. Clontech Labs, Inc.*, 77 USPQ2d 1161 (Fed. Cir. 2005), and under the U.S. Patent and Trademark Office Written Description Training Materials (Revision 1, March 25, 2008).

A Structure/Function Relationship has been established.

The Examiner maintains the rejection for lack of written description on the ground that "applicants have not related the subgenus of structure to the acyclonucleotide incorporation function" (Office Action mailed April 21, 2009, page 4). Appellants respectfully disagree with this assertion. The disclosure of the specification, working examples, and declaratory evidence demonstrates a relationship between the structure recited in the claims and acyclonucleotide incorporation function. The claims require that the DNA polymerase have an amino acid sequence with at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4 (Vent™). The claims also require that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions.

The specification explains that proteins can display sequence similarity over short stretches of primary amino acid sequence (specification, page 14). These patches are thought to occur most often at essential protein interfaces, such as those involved in catalysis, substrate binding, or protein-protein recognition. The degree of sequence similarity, particularly in conserved sequence motifs, is predictive of the degree to which the proteins will behave similarly in both physical properties and catalytic function (specification, page 14, lines 10-16). The claims include just such a motif, by requiring that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions. The sequences of the 15 amino acid motifs and the DNA polymerase in which each is found are shown in Table 3 of the specification at pages 20-21. Each motif is within a conserved region having a role in substrate binding, known as "motif B" as defined by Delarue et al. (*Protein Eng.* 3:461-467, 1990; see citation to Delarue et al. in the specification at page 21 under Table 3; Delarue et al. was submitted with the Information Disclosure Statement filed on May 9, 2002, and is attached as **Exhibit A**). Delarue et al. does not recognize or discuss acyclonucleotide activity of any DNA polymerases. However, Delarue et al. indeed indicates that motif B is involved in DNA polymerase function. In the Discussion section, Delarue et al. states:

> ***From structure to function*. Considerable biochemical evidence points to the imporance of [motifs A, B and C] in the DNA polymerase activity**. A synthesized *E. coli* pol I oligopeptide corresponding to the N-terminal-most two-thirds of the loop region connecting helices O and P (**motif B**-see Figure 4) **has been shown to bind deoxynucleotide triphosphate substrates of pol I** as well as duplex DNA (Mildvan, 1989)" (Delarue et al., page 465, right col., lines 9-15; emphasis added).

This structure/function relationship between motif B and polymerase activity is further confirmed in declaratory evidence submitted during prosecution of the present application. In the Declaration by Dr. William Jack, filed on May 4, 2006 ("the Jack Declaration"; copy attached as **Exhibit B**), it states that Dr. Jack and colleagues have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides and enhanced incorporation with Vent A488L and 9°N 485L DNA polymerase mutants, citing to Gardner et al. (*J. Biol. Chem.* 279(12):11834-11842, 2004; Gardner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Gardner et

al. shows an alignment of Family B DNA polymerases in Figure 1. As is clear from the Figure, the "Region III" active site overlaps with the 15 amino acid motif recited in Appellants' claims. As provided in the Jack Declaration, Gardner discusses the physical basis for incorporation of acyclonucleotides at page 11841. This discussion mentions the A288 residue in Vent™, which is in the active site and in the 15 amino acid motif in Appellants' claims. A relationship between "Region III", containing the 15 amino acid motif, and polymerase function, had been previously noted, e.g., in Hopfner et al. (*Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999; Hopfner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Hopfner et al. reports the crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. Hopfner et al. provide a structure based sequence alignment of archaeal family B polymerases, and show that Region III (which contains the 15 amino acid motif) is in the active site of these enzymes (see Hopfner et al., page 3603, col. 1, Figure 3, and col. 2, third full paragraph). Hopfner et al. discusses the conserved $KX_3NSXYGX_2G$ motif, which is a sub-motif within Appellants' claimed 15 amino acid motif, in the section entitled "Polymerase Active Site", noting that it and a second motif "form the bottom of the nucleotide-binding site" (Hopfner et al., page 3603, right col., lines 31-32). The subgenus of structure (i.e., the 15 amino acid motif) is clearly related to function.

Although there was recognition in the art that conserved motifs found in polymerases are involved in polymerase activity, it is Appellants who recognized and now claim a method of using a specific genus of polymerases which possess acyclonucleotide incorporation function. The set of 15 amino acid motifs specified by SEQ ID NOs 5-22 and recited in the claims are highly related to each other. SEQ ID Nos 6-17 differ from SEQ ID No 5 by three or fewer residues. SEQ ID Nos 18 and 20-22 differ from SEQ ID Nos 5 by six or fewer residues. Motifs of polymerases having sequences sharing less than 30% overall identity with Vent™ (and thus which are outside the scope of the claims) have motifs which differ from SEQ ID No 5 by seven or more residues (see, e.g., SEQ ID Nos 23-30 at page 21, Table 3 of the specification).

A structure/function relationship is not only supported by an understanding of the 15 amino acid motif and its role in enzymatic function. It is also supported by Appellants' working examples. Every DNA polymerase tested that meets the structural requirements of the claims has acyclonucleotide incorporation activity. Indeed, four different DNA polymerases,

Vent™, Deep Vent™, *Pfu*, and 9N™, showed the ability to incorporate acyclonucleotides (specification, Example 6). Two variants of these enzymes, Vent™/A488L, and 9N™/A485L, were also shown in incorporate acyclonucleotides (specification, Example 11). By contrast, Thermosequenase, which is a Taq DNA polymerase variant that lacks the 15 amino acid motif required by the claims, showed a much stronger preference for dideoxyoligonucleotides over acyclonucleotides (specification, Examples 5 and 12). The application therefore establishes the correlation between the sequence motif and the function recited in the claims.

In addition, the Jack Declaration includes an Appendix with data confirming that an archaeon Family B polymerase from *Methanococcus maripaludis*, having 41% sequence identity with Vent DNA polymerase, utilizes acyclonucleotides as a substrate (Jack Declaration Appendix I, attached as **Exhibit B**). Thus, support for a relationship between the DNA polymerases recited in the claims and acyclonucleotide incorporation function has been provided in by information in the specification regarding sequence similarity and function, exemplification of a relationship between the claimed structure and function in the specification, and data and information provided with the Jack Declaration.

The Examiner maintains his rejection without offering any reason *why* the claimed structure/function relationship allegedly has not been established. For example, in the Office Action mailed May 29, 2008, the Examiner said that "[w]hile Applicants comments regarding the homogeneity shared between this group of polymerases continues to be acknowledged, such is acknowledged in light of the degree of the vast majority of DNA polymerases, many of which have a high degree of homogeneity and not all of which share the ability to incorporate acyclonucleotides into a DNA fragment" (Office Action mailed May 29, 2008, page 4). Appellants have related specific structural features (overall sequence identity and the presence of a 15 amino acid motif in the active site of the enzyme) to function (acyclonucleotide incorporation function). The Examiner has provide no *reason* to doubt Appellants correlation. The Examiner is not entitled to substitute his personal skepticism for statements and evidence provided by the Appellants.

Written description support for the claims is met under Invitrogen Corp. v. Clontech Labs, Inc. 77 USPQ2d 1161 (Fed. Cir. 2005).

Relevant legal precedent also confirms that the written description requirement is satisfied for the present claims in view of the present specification. The decision in *Invitrogen Corp. v. Clontech Labs, Inc.* 77 USPQ2d 1161 (Fed. Cir. 2005) requires a finding that the claims are adequately described. To emphasize this point, Appellants reiterate a close comparison between *Invitrogen* and the present claims here. The claim at issue in *Invitrogen* read:

> 1. An isolated polypeptide having DNA polymerase activity and substantially reduced RNAse H activity, wherein said polypeptide is encoded by a modified reverse transcriptase nucleotide sequence that encodes a modified amino acid sequence resulting in said polypeptide having substantially reduced RNase H activity, and wherein said nucleotide sequence is derived from an organism selected from the groups consisting of a retrovirus, yeast, Neurospora, Drosophila, primates and rodents.

The specification supporting the claim had only a single example of a polymerase having the recited activity. The court found that the claim met the written description requirement because, (1) at the time of the invention, sequences of reverse transcriptase (RT) genes were known; (2) members of the RT gene family shared significant homologies from one species to another; (3) the written description taught that the invention can be applied to RT genes of other retroviruses; and (4) the specification cited references providing the known nucleotide sequences of those genes.

It must be noted that, unlike the claim in *Invitrogen* which recites no structural limitations, the pending claims include explicit recitation of structural features (overall homology and a 15 amino acid motif). The present specification provides six specific examples of DNA polymerases that fall within the claims. As for the other factors from *Invitrogen*, (1) sequences of many DNA polymerases were known when the present application was filed; and (2) members of the DNA polymerase gene family share significant homologies from one species to another. See the present specification, e.g., at page 3, lines 8-21; and page 10, line 12, to page 15, line 34. For (3), the written description of the present case clearly teaches that the invention can be applied to DNA polymerases other than the ones specifically exemplified. See, for example, page 19, lines 15-27, which teaches:

> The similarity of incorporation patterns with these selected enzymes suggests that not only these archaeon DNA polymerases, but a larger family of DNA polymerases could share the ability to incorporate acyclo to a greater extent than dideoxy terminators. Since *Pfu*, Deep Vent® and 9°N™ DNA polymerases have greater than about 70% sequence identity with Vent DNA polymerase, other enzymes with equivalent or greater identity can reasonably be expected to perform as Vent® (exo-) DNA polymerase in this invention. Notably, those enzymes for which no significant sequence similarity is found (i.e., Family A DNA polymerases such as Taq) do not perform in similar ways in the current invention. This fact leads us to believe that archaeon enzymes showing intermediate identity, namely those between about 20 and 70% identity are reasonably candidates for this invention.

As to (4), the specification cites references providing the known sequences of such other DNA polymerases (see, for example, page 10, line 22; page 14, line 18; page 14, line 19; page 15, lines 19-24). Moreover, the sequences of other DNA polymerases are known and need not be fully presented in the specification to satisfy the written description requirement. See *Capon*, 418 F.3d at 1358.

Appellants maintain that, with regard to every relevant fact relied upon by the court, the present case has at least as much, or more description than was provided in *Invitrogen*.

The Examiner disputes this point because the claims encompass incorporation of acyclonucleotides into DNA and

> [t]his is not a property of a DNA polymerase that is well known in the art and the applicants have not adequately described this supposedly new function of a specific sub-genus of DNA polymerases. This is in contrast to the claims of Invitrogen in which the homologies of the encompassed DNA polymerases were high and that region responsible for reduced RNAse H activity in each of these DNA polymerases known such that the encompassed DNA polymerase variants known. (Office Action mailed April 21, 2009, page 5).

Appellants explain in detail the relationship between structure provided and acyclonucleotide function above. As discussed, Appellants have demonstrated (through several examples) that DNA polymerases that do have the claimed sequence do have the recited activities, and a DNA polymerase that does not have the claimed sequence does not have the recited activity.

Moreover, the fact that the present claims recite use to perform a newly discovered function (incorporation of acyclonucleotides) does not distinguish the present case from *Invitrogen*, as asserted by the Examiner. The claims in *Invitrogen* also related to DNA polymerases that have a new function (reduced RNAse H activity). The Examiner is correct that the *region* of DNA polymerase sequence that was responsible for RNAse H activity was previously known. As discussed above, the relevant region of DNA polymerases (region III) involved in the present claims was also known (and known to be important for activity, just not for this activity). The present specification demonstrates that this known region is important for a new activity, much like the specification in *Invitrogen* demonstrated that changes in a known region could reduce activity. Closer factual scenerios in fact would be difficult to find!

Furthermore, Appellants fail to see how acyclonucleotide function of the DNA polymerases renders this case distinguishable from *Invitrogen*. In that case, a single example of an enzyme having a desired function (reduced RNAse H activity) was adequate to support the claims.

Appellants' recognition of a class of polymerases which incorporate acyclonucleotides is new, and Appellants have linked the functional activity with structure and a characterized structural, functional motif (i.e., the 15 amino acid motif). There is no basis for distinguishing the present case from *Invitrogen*. The Examiner suggests that *Invitrogen* is not applicable because "the homologies of the encompassed DNA polymerases were high." Yet the *Invitrogen* claim is completely devoid of structural limitations, and recites polymerases from organisms as diverse as viruses, yeasts, and primates! If unspecified sequences from such varied species have "high" homology in the Examiner's view, Appellants fail to understand how homologies between sequences encompassed by the present claims, which recite concrete structural limitations, are not also "high."

In a further attempt to distinguish *Invitrogen*, the Examiner stated that

> the description held by Invitrogen is specific to the claims of invitrogen [sic], based upon the specification and art as well as a. Actual reduction to practice, b. Disclosure of drawings or structural chemical formulas, c. Sufficient relevant identifying characteristics, such as: Complete structure, ii. Partial structure, iii. Physical and/or chemical properties, iv. Functional characteristics when coupled with a known or disclosed correlation between function and structure, d. Method

of making the claimed invention, e. Level of skill and knowledge in the art and f. Predictability in the art. (Office Action mailed April 21, 2009, carryover paragraph from pages 5-6).

Legal decisions would be meaningless as precedent if they could be applied only to a single set of facts. Appellants have provided a close comparison of (i) the facts in the *Invitrogen* case and the (stronger) facts here; and (ii) the claims of the present application and a claim from *Invitrogen* for which written description was affirmed. There has been no showing that *Invitrogen's* claimed genus all had "high" homology or "known" function such that the present claims can be distinguished from the case. No other bases for finding *Invitrogen* inapplicable have been offered.

Written description support for the claims is met under the U.S. Patent and Trademark Office Written Description Training Materials

The Examiner compared the present claims to the U.S. Patent and Trademark Office Written Description Training Materials (hereinafter, the "Guidelines") and found lack of description in Appellants' claims compared to claim 2 in Example 11 of the Guidelines because "claim 2 is drawn to a nucleic acid having 85% identity to a specific sequence, a partial structure. This is relative to the instant claims which require even less partial structure of 30% identity." (Office Action mailed April 21, 2009, page 7).

The Examiner has not analyzed Appellants' claims in view of the knowledge of DNA polymerase structure and the requirement of a conserved motif which is associated with enzymatic function. Claim 2 in Example 11 of the Guidelines concerns a claim to nucleic acid encoding hypothetical polypeptide having "activity X". In contrast to the present claims, the hypothetical polypeptide encoded by the nucleic acid does not share significant sequence identity with any known polypeptide or polypeptide family. Also unlike the present claims, the specification for this hypothetical example discloses only a single nucleic acid sequence that encodes a polypeptide having "activity X". Any comparison of the present claims to Example 11 should take these facts into consideration. Another important factor for analysis in Example 11

is the presence of a disclosed or art-recognized correlation between structure and function. Appellants have provided this correlation.

Example 5 of the Guidelines presents a fact pattern much more analogous to Appellants claims, and is a more appropriate basis for comparison. Example 5 concerns a claim to an "isolated protein comprising Protein A," wherein Protein A includes the amino acid sequence of SEQ ID NO:1, has the ability to bind and activate Protein X, and is purified by a recited set of conditions. The sequence of SEQ ID NO:1 in this hypothetical claim has 10 amino acids. Likewise, Appellants' claims recite DNA polymerases that include a 15 amino acid motif and have a specific binding and activity function, which is the ability to incorporate acyclonucleotides in a polymerase extension reaction. The polymerases are not defined by purification conditions. However, significant structural definition for the polymerases is provided by requiring at least 30% identity to SEQ ID NO:4.

In the hypothetical fact pattern set forth for Example 5, claim 1, the specification fails to disclose the complete structure of Protein A and it fails to disclose any art recognized correlation between the structure of the claimed protein and its function of binding and activating Protein X. Nonetheless, written description is affirmed for the claim because the specification discloses a partial (10 amino acid) sequence of Protein A and because relevant identifying characteristics are provided in the form of its ability to bind and activate Protein X, and purification features.

If anything, the present specification provides more description support for the claims than is provided for claim 1 of Example 5 of the Guidelines. Appellants' specification describes examples of complete structures for polymerases that fall within the claims. Appellants' 15 amino acid motif imposes greater structural definition for a polymerase than the 10 amino acid sequence defining the hypothetical polypeptide of Example 5. Appellants' polymerases possess a binding ability and activity (acyclonucleotide incorporation) which is just as well defined as those of the hypothetical polypeptide of Example 5. Whereas no correlation of protein structure with function is provided in Example 5, Appellants' provide detailed structure/function correlation, as set forth above. In this aspect, Appellants provide more support than the Guidelines require. Another factor favoring support for the hypothetical polypeptide was the specification's disclosure of methods for isolating the polypeptide and a working example showing the polypeptide was successfully isolated. Appellants' have also shown that one of skill

in the art can make and use polypeptides as claimed, and that polypeptides have the recited function.

Claim 33 is not invalid for lack of written description

Claim 33 stands rejected for lack of written description. Claim 33 specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the genus of polymerases encompassed by the claim is smaller than that of claim 32. Thus, the level of description required is reduced as compared with claim 32. Appellants' specification demonstrates that multiple polymerases within the genus possess acyclonucleotide function. (Appellants emphasize that polymerases from the broader genus have this function as well; see the Jack Declaration, Appendix I, which shows that a *Methanococcus* DNA polymerase having only 41% sequence identity to Vent™ DNA polymerase incorporates acyclonucleotides more efficiently than dideoxynucleotides.) Even if claim 32 were not fully supported by the specification (which Appellants do not concede), claim 33 would be.

Claim 34 is not invalid for lack of written description

Claim 34 stands rejected for lack of written description. Claim 34 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. Given the further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions), the genus of polymerases encompassed by the claim is smaller than that of claim 33. The level of description required for this claim is reduced as compared with claim 32. Even if claim 32 were not fully supported by the specification, claim 34 would be.

Claim 35 is not invalid for lack of written description

Claim 35 stands rejected for lack of written description. Claim 35 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. Because it covers fewer motifs, this

claim refers to a genus of polymerases that is smaller than that encompassed by claim 32. The level of description required to support this claim is less than required for claim 32.

Claim 36 is not invalid for lack of written description

Claim 36 stands rejected for lack of written description. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-17. The genus of polymerases encompassed by this claim is even smaller than that of claim 32 and requires less description to be adequately supported.

Claim 37 is not invalid for lack of written description

Claim 37 stands rejected for lack of written description. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

Claim 38 is not invalid for lack of written description

Claim 38 stands rejected for lack of written description. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

Claim 40 is not invalid for lack of written description

Claim 40 stands rejected for lack of written description. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

Claim 41 is not invalid for lack of written description

Claim 41 stands rejected for lack of written description. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

Claim 42 is not invalid for lack of written description

Claim 42 stands rejected for lack of written description. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

In conclusion, the provided teachings in the specification, examples, sequences, declaratory evidence, and data are more than sufficient to describe function and support description of the claims. The Examiner has not established otherwise. For reasons set forth above, withdrawal of the rejection of claims 32 and 39 as allegedly lacking written description is respectfully requested.

Ground of Rejection 2:

Claims 32 and 39 are not invalid for lack of enablement

Pending claims 32 and 39 stand rejected for lack of enablement. The Examiner states that the specification, while being enabling for a method comprising providing a DNA polymerase selected from the group consisting of Vent™, Deep Vent™, *Pfu*, and 9°TM or the specifically disclosed variants of claim 43, "does not reasonably provide enablement for any method comprising providing a DNA Polymerase having an amino acid sequence that shows a mere 30% overall identity with that of SEQ ID NO:4 and further includes a 15 amino-acid motif that is identical to SEQ ID NO:5 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO..." (Office Action mailed April 21, 2009, pages 8-9). The Examiner stated that "determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue" (Final Office Action mailed April 21, 2009, pages 11). Appellants have previously reviewed the factors set forth in *In re Wands* (858 F.2d 731, 8 USPQ2nd 1400, Fed. Cir. 1988) with respect to the present claims and review them here, in response to the Examiner's assertion that Appellants' burden has not been met.

First, Appellants address the Examiner's comments regarding *Wands* factor (2), Amount of Direction or Guidance. In the Final Office Action mailed April 21, 2009, the Examiner maintained that guidance was lacking as to DNA polymerases which have the ability to incorporate acyclonucleotides into a DNA template, and requested clarification as to how the 15 amino acid motif correlates with acyclonucleotide function (Office Action mailed April 21, 2009, page 12). As explained above in the arguments for written description, the 15 amino acid motif is a highly conserved motif in the active site of family B DNA polymerases which plays a role in substrate binding. The Examiner disputes a structure/function correlation because "applicants have not disclosed such a single motif but rather continue to refer to any of a number of motifs or variants thereof"(Office Action mailed April 21, 2009, page 12). Some variability within the genus of motifs is permitted, given that variable polymerases share acyclonucleotide incorporation function. For example, both 9°N polymerase and Vent™ incorporate

acyclonucleotides, although their 15 amino acid motifs differ by three amino acids (compare SEQ ID NO 5 and SEQ ID NO 7 at page 20, Table 3 of the specification). A *Methanococcus maripaludis* DNA polymerase having a more divergent sequence also possesses acyclonucleotide incorporation activity. The claims do require a degree of conservation of sequence, which is clearly expressed in the claims. The fact that variable polymerases share a specific function does not render them "unpredictable."

As to the (1) Quantity of Experimentation Necessary, and (3) Presence or Absence of Working Examples, Appellants reiterate that one or ordinary skill could make and test all polypeptides within the scope of the claims to determine their ability to extend a DNA primer or incorporate acyclonucleotides (including to determine their ability to preferentially select acyclonucleotides). Appellants' working examples include demonstration of activity of multiple species of DNA polymerases set forth in the specification and in declaratory evidence discussed herein.

As to (5) State of the Prior Art, and (7) Predictability of the Art, Appellants note, and the Examiner has acknowledged, that the prior art with regard to DNA polymerases and their classification is extensive. However, Appellants disagree with the Examiner's assertion that "determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue" (Office Action mailed April 21, 2009, page 11). Appellants have identified polymerases which have acyclonucleotide incorporation function by virtue of structural and physical characteristics distinctive of well-characterized DNA polymerases. These characteristics include overall sequence identity to a polymerase, and the presence of a conserved motif. Appellants have shown that all members tested within the genus of polymerases have the recited activity. The Examiner's only discernible reason for declaring these features unpredictable is the breadth of the genus of polymerases. This improperly disregards Appellants' demonstration of activity for multiple species. It also disregards the fit of Appellants' observed activity with well characterized classification schemes for DNA polymerases (among which one finds substantial variability despite conserved nucleotide polymerase activity).

As to the (4) Nature of the Invention and (8) Breadth of the Claims, Appellants reiterate that DNA extension reactions are well within the skill of those of ordinary skill. As part of their invention, Appellants have described a class of DNA polymerases that can incorporate acyclonucleotides, and have shown function for six different species within the class. Given the demonstrated correlation of structure with function and other reasons provided above, Appellants disagree with the Examiner's assertions that the scope of the claims is not enabled.

As to (6) Relative Skill of those in the Art, Appellants submit, and the Examiner has agreed, that the relative skill of those in the art is very high.

Claim 33 is not invalid for lack of enablement

Claim 33 stands rejected for lack of enablement. Claim 33 depends from claim 32 and specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the breadth of the claim is smaller than that of claim 32. The scope of enablement provided by the disclosure is more than sufficient to support the scope of this claim, not least because multiple polymerases that fall within the claimed genus are exemplified.

Claim 34 is not invalid for lack of enablement

Claim 34 stands rejected for lack of enablement. Claim 34 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. This further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions) provides a claim of smaller breadth than claim 32 and which is more than supported by the disclosure.

Claim 35 is not invalid for lack of enablement

Claim 35 stands rejected for lack of enablement. Claim 35 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it

contains up to 3 amino acid substitutions as compared with the SEQ ID NO. This claim covers fewer motifs than claim 32 and is enabled for its full scope.

Claim 36 is not invalid for lack of enablement

Claim 36 stands rejected for lack of enablement. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-17. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 and is enabled by the disclosure provided.

Claim 37 is not invalid for lack of enablement

Claim 37 stands rejected for lack of enablement. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. The genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

Claim 38 is not invalid for lack of enablement

Claim 38 stands rejected for lack of enablement. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

Claim 40 is not invalid for lack of enablement

Claim 40 stands rejected for lack of enablement. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

Claim 41 is not invalid for lack of enablement

Claim 41 stands rejected for lack of enablement. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

Claim 42 is not invalid for lack of enablement

Claim 42 stands rejected for lack of enablement. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is enabled for its full scope.

In light of the above, Appellants submit that claims 32 -42 satisfy the enablement requirement. Allowance of the claims is requested.

Respectfully submitted,

Date: February 2, 2010

/Margo H. Furman/
Margo H. Furman, PhD, JD
Registration Number 59,812
Tel. No. (617) 248-4073

PATENT GROUP
CHOATE, HALL & STEWART LLP
Two International Place
Boston, MA 02110
Tel: (617) 248-5000
Fax: (617) 502-5002
patentdocket@choate.com

CLAIMS APPENDIX

1-31. (Canceled)

32. (Previously presented) A method comprising steps of:

providing a DNA polymerase having an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO;

contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and

incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides.

33. (Previously presented) The method of claim 32, wherein the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4.

34. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-22.

35. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-17 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.

36. (Previously presented) The method of claim 35, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-17.

37. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.

38. (Previously presented) The method of claim 37, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8.

39. (Previously presented) The method of claim 32 or 33, wherein the step of incubating comprises incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides, and preferentially incorporates acyclonucleotides.

40. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22.

41. (Previously presented) The method of claim 35, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17.

42. (Previously presented) The method of claim 37, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-8.

43. (Previously presented) The method of claim 32 or 33 wherein the DNA polymerase is Vent™, Deep Vent™, 9°N, *Pfu*, Vent™/488L, or 9°N/485L.

EVIDENCE APPENDIX

Appellants had provided the following evidence during prosecution of the instant application:

Exhibit A: Delarue et al., *Protein Eng.* 3:461-467, 1990. This reference was cited in the Information Disclosure Statement and Form PTO-1449 filed on May 9, 2002, and was entered into the record on May 13, 2002. The Form PTO-1449 was initialed by the Examiner on September 29, 2004, confirming that the reference was entered into the record.

Delarue et al. is attached hereto at pages 31-37.

Exhibit B: Declaration of William Jack, accompanying references and Appendix I. The Declaration was submitted with four references, listed below, and Appendix I along with a response to Office Action filed May 4, 2006, and was entered into the record in PAIR on May 9, 2006 as the entry designated "Rule 130, 131 or 132 Affidavits." Entrance into the record was confirmed by the Examiner's reference to this Declaration on page 3 of the Advisory Action mailed on July 5, 2006.

The Declaration of William Jack is attached hereto at pages 38-45.

Rodriguez et al., *J. Mol. Biol.* 299:447-462, 2000, is attached hereto at pages 46-61.

Gardner et al., *J. Biol. Chem.* 279(12): 11834-11842, 2004, is attached hereto at pages 62-70.

Hashimoto et al., *J. Mol. Biol.* 306:469-477, 2001, is attached hereto at pages 71-79.

Zhao et al., *Structure* 7(10):1189-1199, 1999, is attached hereto at pages 80-90.

Hopfner et al., *Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999, is attached hereto at pages 91-96.

Appendix I is attached hereto at pages 97-99.

RELATED PROCEEDINGS APPENDIX

Not applicable.

EXHIBIT A

Protein Engineering vol.3 no.6 pp.483–487, 1990

An attempt to unify the structure of polymerases

Marc Delarue, Olivier Poch[1], Noel Tordo[1], Dino Moras and Patrick Argos[2]

Department of Crystallography and Biophysics, IBMC du CNRS, 15 rue René Descartes, 67084 Strasbourg, [1]Institut Mérieux, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France and [2]European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 6900 Heidelberg, FRG

With the great availability of sequences from RNA- and DNA-dependent RNA and DNA polymerases, it has become possible to delineate a few highly conserved regions for various polymerase types. In this work a RNA polymerase sequence from bacteriophage SP02 was found to be homologous to the polymerase domain of the Klenow fragment of polymerase I from *Escherichia coli*, which is known to be closely related to those from *Staphylococcus pneumoniae*, *Thermus aquaticus* and bacteriophages T7 and T5. The alignment of the SP02 polymerase with the other five sequences satisfactorily recovered the conserved motifs in these proteins. Three of the motifs matched reasonably all the conserved motifs of another DNA polymerase type, characterized by human polymerase α. It is also possible to find these three motifs in mammalian DNA-dependent RNA polymerases and two of them in RNA polymerase I and RNA terminal transferases. These latter two motifs also matches two of the four motifs recently identified in all RNA-dependent polymerases. From the known tertiary structure of the Klenow fragment of *E. coli* pol I, a spatial arrangement can be inspired for these motifs. In addition, numerous biochemical experiments suggesting a role for the motifs in a common functions (dNTP binding) also support these inferences. This speculative hypothesis, attempting to unify polymerase structure at least locally, if not globally, under the pol I fold, should provide a useful model to direct mutagenesis experiments to probe template and substrate specificity in polymerases.

Key words: catalytic domain/DNA polymerase/RNA polymerase/sequence/structure

Introduction

The number of available protein sequences is growing rapidly due to the facility of nucleotide sequencing techniques. One particular class evolved and augmented at the polymerase levels which is central to the duplication and expression of genes. Polymerases can use RNA or DNA as a template (RNA- or DNA-dependent); the product can also be RNA or DNA. Polymerases are found both in eukaryotes and prokaryotes, though surprisingly efforts have often concentrated on those from viruses. One way to use the information contained in all these sequences is to try to align them and thereby discover their strongest internal features and similarities. This has been achieved for DNA-dependent DNA polymerases, where three main subfamilies have been identified. One of them contains the Klenow fragment of *Escherichia coli* polymerase I, whose three-dimensional structure is known (Ollis *et al.*, 1985b), and polymerases from phages T5 (Ollis *et al.*, 1985b; Argos *et al.*, 1986) and T3 (Hanvit and Ito, 1989), and from *Thermus aquaticus* (Lawyer *et al.*, 1989) and *Staphylococcus pneumoniae* (Lopez *et al.*, 1989). This family will be referred to as the pol I family. For another set of DNA-dependent DNA polymerases (Wang *et al.*, 1989), of those homologous to the human polymerase α (hereafter referred to as pol α), more than 10 sequences from various species are known. A third subfamily of DNA-dependent DNA polymerases, homologs called the pol B type, has only two members: DNA polymerase II (Matsukage *et al.*, 1987) and terminal transferase (Peterson *et al.*, 1985; Kotani, 1986; Zmudzka *et al.*, 1988). Until now, just one DNA-dependent DNA polymerase sequence, from the SPO1 bacteriophage (Rabin and Romig, 1986; Jong *et al.*, 1987), resisted alignment with any of the three aforementioned types.

Clearly, the aim of these alignments, apart from evolutionary implications, is the identification of the regions essential for polymerase function, since these sequence segments should appear in the most conserved. Generally, a good number of aligned sequences will ensure sufficient variability to identify the functionally required regions. For the pol I type, the five previously aligned (Ollis *et al.*, 1985b; Argos *et al.*, 1986; Lawyer and Ito, 1989; Lopez *et al.*, 1989) sequences are sufficiently close so as not to allow confident delineation of the absolutely required motifs.

In the present work, it has been found that the polymerase from bacteriophage SPO2 can be aligned, using a sensitive method, with the polymerase portion of the Klenow fragment in the C-terminal part of the protein [the N-terminal domain has a 3′-5′ exonuclease function (Freemont *et al.*, 1988)]. The total alignments of the C-terminal part of the six proteins of the pol I type is presented. The relationship of SPO2 polymerase with those from phages T7 and T5, *Lymnaeomicrobium*, *Thermus* and *E.coli* is sufficiently distant that highly conserved regions can now be referred to five in number. Interestingly, three of the five regions match reasonably with the three most conserved motifs of DNA-dependent DNA pol α, suggesting that the two polymerase types may share a common tertiary fold, or at least contain similar local tertiary architecture required for similar functions. These motifs are likely to represent modules required for the polymerase structure and activity.

Searches were then performed to detect such sequence patterns in other polymerase families. All three motifs could be found in DNA-dependent RNA polymerases that consist of only one subunit (see Moreno *et al.*, 1987); two motifs were found in pol β, in the same linear arrangement and maintaining the strictly conserved residues. In addition to this, an examination of several aligned RNA-dependent RNA polymerases as well as reverse transcriptases, for which four highly conserved motifs have been highlighted (Kamer and Argos, 1984; Poch *et al.*, 1989), all lent the suggestion that two of these motifs could match the two motifs shared by DNA pol B, pol α and pol β. These sequence similarities are further supported by a consistently significant alignment between the entire polymerase domain of two members of these different families, namely a DNA-dependent RNA

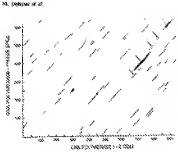


Fig. 1. Homology search curve (degen. 1067) between the Klenow fragment of DNA pol I from *E.coli* and DNA polymerase from bacteriophage RB49. The scores, relative lengths ranged from 2 or 20 to more of 2. Each search point was plotted over the entire relative length, with the largest score designating other smaller scores from the matching/gap searches. The peak values (D) are scored as a measure of standard deviations above the random mean for each relative length. Thus these values in the 4 to 5 or 6 to 7 and thick lines refer to 4.5 or 6.5 to 6.6σ. The data used as a basis to affect the alignments of Figure 2 is indicated by arrows. The remaining alignments shown in Figure 2 are based on conservative searches of all the sequences.

polymerase and an RNA dependent DNA polymerase. The alignment of the C-terminus core motif, containing two acidic groups, is in agreement with those suggested by Argos (1988b) amongst RNA-dependent RNA polymerases, reverse transcriptases and DNA-dependent DNA pol αα.

An examination of these conserved regions relative to the known Klenow fragment tertiary architecture positions them in zones likely to interact with nucleotides. In fact, many biochemical data (Joyce and Steitz, 1987; Lecler *et al.*, 1987b) point to the importance of these motifs in the catalytic process, in the different families of polymerases. The design of site-directed mutagenesis experiments around broadly from the model suggested here, replacing possible common tertiary segments for various polymerases, if not a consensus topology.

Results

Pairwise sequence comparisons were done by a search procedure based on similar characteristics (Argos, 1987). The resulting alignments of the C-terminal polymerase portions of the Klenow fragment of DNA-dependent DNA polymerase I from *E.coli* and polymerases from bacteriophages T7, T5, *S.pneumoniae* and *T.aquaticus* are similar to the ones given by Argos *et al.* (1986a), Lazaro and Bu'Lock (1989), Lopez *et al.* (1989) and Lawyer *et al.* (1989) respectively. The DNA polymerase sequence from bacteriophage RB69 was also compared to the *E.coli*, *S.pneumoniae*, T3 and T7 polymerases. The search matrix and

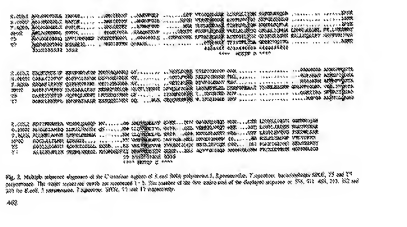
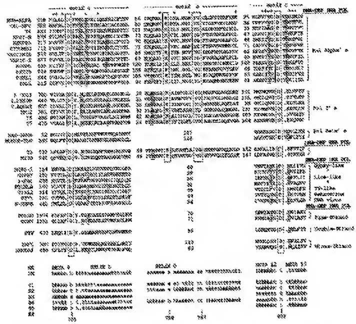


Fig. 2. Multiple sequence alignment of the C-terminus region of *E.coli* DNA polymerase I, *Epstein-Barr*, *T.aquaticus*, bacteriophages RB69, T5 and T7 polymerases. The major conserved motifs are numbered 1–3. The number of the first amino acid of the displayed sequence in T5, T7, 405, 373, 832 and 928 for *E.coli*, *S.pneumoniae*, *T.aquaticus*, RB69, T3 and T5 respectively.

[illegible]



[illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4603542v1

events it is retained. Dayhoff matrix as used by Gribskov and Burgess (1986). A threshold of 60% conservation hides five regions of high homology, which are indicated in Figure 2. These regions are characterized by the consensual sequences (F)(pivv)xx(pa)(D)(L)xxx(L, in the first region (* refers to a given alignment position occupied by any amino acid; (X) indicates an amino acid shared symmetrically conserved by the six sequences, and a single letter refers to exact conservation in all the sequences); (T)(D)F for region 2, DxG(pa)xL for region 3, K*****(YW) for region 4 and VH(EQ)G for region 5.

A multiple alignment was also built for 12 sequences of DNA-dependent DNA polymerases on three main, yeast and several viruses. These sequences are essentially the ones generated by Wang et al. (1989) together with two recent additions to the nrDB sequence data bank: toxic polymerase from another adenovirus, Wrair#2, and one from a nuclear polyhedrosis virus, Spodoptera. The resulting conservation profile using a window of five residues left three regions above a 60% conservation threshold. They are characterized by DxwSh(LY)(FG)(L)***(EK)xx(FYG) and (Y)(N)(LIVM). as previously noted by other authors (Bernad et al., 1987). A comparison of the exact conserved sequences for the pol I and pol α types suggests an overlap of these three regions (I and pol α) with regions 3, 4 and 5 of the pol I type; these regions will hereafter be called motifs A, B and C (Figure 3). These three motifs are common on aminoacid residues motif A contains a strictly conserved aspartate at the junction of a beta strand and an alpha helix, motif B contains an alpha helix with positive charges and motif C has a district of negative charges location in a beta–turn–beta secondary structure, as implied from the Klenow structure (Figure 4). Great variability in the lengths separating the different conserved regions was observed for the pol α: 90–120 residues between motifs A and B (average 98) but 25–150 residues between motifs B and C (average 68). In the pol I, these values correspond respectively with 30 and 100 residues, on average.

A set DNA-dependent DNA polymerase β (Mossbaugh et al., 1987) can be easily aligned with a known tertiary structure, the bacteriophage T4 enzyme (Wang et al., 1989). Though many residues are identically conserved amongst the two sequences, an N-terminal and a C-terminal fragment may be related to motifs A and C of DNA polymerases (Figure 3).

For DNA-dependent RNA polymerases, the three motifs could only be found in the putative mode of only one subunit. Four different sequences of this type are known, the two most distantly related being from phage T3 and yeast mitochondria (Masters et al., 1987). The distances between motifs computed well with the ones of the pol I type. The sequence fragments corresponding to the motifs are presented in Figure 3, together with those found in DNA, pol αc, pol β and pol Rb.

Secondary structure predictions averaged over each set of sequences (DNA pol c, pol I and pol β) and pol DNA-dependent RNA pols) were made according to the procedures of Argos (1985) and are given in Figure 3. Except for pol β, for which only two structures are known, they often show agreement amongst themselves as well as with the pol I structure, especially for motifs A and C.

Recently, Poch et al. (1989) identified four motifs shared by all RNA-dependent RNA and DNA polymerases (24 different sequences). Two of these motifs were found to be similar to two of those contained in the DNA-dependent polymerases, namely, motifs A and C. These two motifs contain the strictly conserved residues of the same name mentioned. The tertiary structure predictions agree with the observed structure in the corresponding

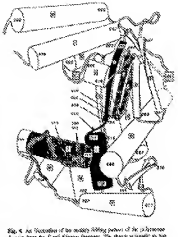


Fig. 4. An illustration of the motifs sitting pattern of the β-hairpin domain from the *E.coli* Klenow fragment. The thumb is mostly as has given in Ollis et al. (1988). The four α-helix thumb region, in ascending from the N-terminus to the C-terminus, corresponds respectively to motifs A (dark region), B (narrow dotted) and C (horizontal stripes), whose sequences are given in Figure 3. Negative positive residues in the secondary structures are mentioned. Polypeptide are numbered consecutively and α-helices are designated by letters according to the alphabet. The structural position of the conserved lysine active is indicated in each of the motifs

regions of the Klenow fragment (Poch et al., 1989). Recurrent fragments from 12 monomeric one DNA-dependent polymerases (1.e. from each of the subfamilies: enzyme transcriptases, photosystem- and double-strand DNA polymerase) are shown aligned in those matched in the DNA pols (Figure 3).

Several total RNA-dependent DNA polymerases and reverse transcriptase sequences were then compared to those from pol Ia and pol αc using the same sequence comparison techniques as above (Argos, 1987). Although the different polymerases comparison, extensive sequence similarities were found between the total hepatitis B RNA-dependent DNA polymerase from Woodchuck (Seilhem et al., 1987) and Herpes simplex virus DNA polymerase, a member of the DNA pol α family (Gibbs et al., 1985). The motifs as shown in Figure 5 and the resultant alignment in Figure 6. It is noteworthy that 31% of the aligned residues are identical. In this alignment, two conserved regions can be delineated that correspond to motifs A and C. The homology between these two sequences provides a possible link between the DNA- and RNA-dependent polymerases.

Discussion

DNA-dependent DNA polymerases

Sequence similarities. In this work, a new member was added to the pol I type of DNA-dependent DNA polymerases. An analysis of the multiple alignments for the pol I types as well as for the pol α type pointed to three conserved motifs, built as

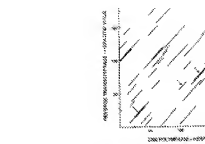


Fig. 3. Homology search chart. 1983 human hexokinase sequence is fitted sequentially to each sequence in the 1984 dataset of three polymerases. The search window length ranged from 9 to 36 in steps of 3; the search peaks are plotted over the same window length, with the longest value connecting each search point to the midpoint. The peak values fill an annulus as a number of isolated correlations are found for nearly every 3-mer search target. The first letters J for 3 × 3-mer, plus Ser, Ala × 3, C × 3, Gly, Ala × 3, C × 3, Ala, and Arg of Ser, Ala × 3, C × 3, Ala. The path used as a basis to align the sequences shown in Figure 4 is indicated by arrows.

conserved. The distances between these motifs are quite variable. However, within each polymerase type, there is already great variability in the distances between conserved regions. Moreover, it would seem unlikely that those of the most conserved sequence regions in the polymerase domains of pol I or Virgata 21 could be aligned with the three most conserved regions of DNA pol α purely by chance. A different piece of evidence pointing to a possible link between the pol I and the pol α types comes from sequence similarities within the extracellular domain of the T5 polycentre and pol α1 (Cazzolli and Ito, 1989). It is thus reasonable to expect that pol I polymerase domains to be related also. In fact, this observation has been recently extended to all extracellular domains of pol α1 and pol Ia, including SPO2 (Bernad *et al.*, 1989). In addition, the secondary structure predictions, especially for motifs A and C, point to sequence-related structures for pol α1 that correspond with those observed in the pol I Klenow structure.

From the sequences to structure. The three motifs A, B and C are clustered and conservative in the C-terminal fraction of the Klenow fragment and correspond to structural features likely to interact with the DNA (Ollis *et al.*, 1985). Figure 4 highlights the portions of the E coli pol I tertiary structure corresponding to the three diagnostic motifs. Motif A, characterized by a conserved Asp and Ser, corresponds to strand D and helix L. Motif B, with conserved Lys, Tyr and Gly encompasses most of helix O and the following long loop region. Motif C delineates strands I2 and I3 with the mostly conserved negative charges contained in the connecting loop of the β-hairpin. Since the N-terminal Asp of this doublet is the only one universally conserved, it is presumed to be catalytically most important. It is clear that the structural segments A and B are likely to come into contact with DNA and that the strand I2–I3 hairpin loop could place at least one catalytic residue in the polymerase active site. Model building of DNA into the Klenow fragment structure also supports the importance of these regions (Ollis *et al.*, 1985a,b; Warwicker *et al.*, 1985). Furthermore, the C_α atom of motif C Asp882 is within 3.5 Å of that from Asp705, conserved in motif A (E_{coli} residue numbers given in Brookhaven database (Bernstein *et al.*, 1977, file 8DPE). Their spatial proximity would allow both to participate in catalysis. Region 2 of pol β (sequence TCR between strands 7 and 8—see Figure 2 and 4) is also located in the vicinity of this area but no sequence homology with the pol α type could be found for this region. Region 1 fails in an undefined part of the electron density map (see Ollis *et al.*, 1985a, Figure 4).

From structure to function. Considerable biochemical evidence points to the importance of these three motifs in the DNA polymerase activity. A synthesized E coli pol I oligopeptide corresponding to the N-terminal most two-thirds of the loop region connecting helices I3 and I' (motif B—see Figure 4) has been shown to bind strongly to calf thymus terminal deoxynucleotidyl transferase (pol I as well as oxytax DNA (Modak, 1986). Furthermore, photo-affinity labelling with 8-azido-dATP identifies Tyr766 as a residue in the active site (Joyce and Steitz, 1987) while Lys758, also part of this motif, succumbs to chemical labelling. Site-specific using pyridoxal phosphate (Basu and Modak, 1987). A further synthesized peptide (Bhattacharyya *et al.*, 1988; Modak, 1988) corresponding to helix G and strands I2 and I3 (motif C) was not found to bind the pol I substrate, although it spontaneously entered its proper folding; however, this result does not exclude that this peptide, although unable to bind dNTP on its own, can co-operate with other regions of the native protein and be just part of the binding site. In fact, mutant of pol I, which is in the loop folding strands I2 and I3 and sequentially adjacent to the conserved Asp882 of motif C, has been shown to be involved in the binding of [^{32}P]dTTP (Pandey *et al.*, 1987). In addition, the beginning of helix Q, which is close in space to the end of helix Q (motif B) and results A and C, can also be labeled by photoaffinity (Modak *et al.*, 1988).

Similarly, for the pol α type, genetic studies of the Herpes simplex virus polymerase (Larder *et al.*, 1987a) revealed that four out of the six identified drug-resistant polymerase mutants cluster in motifs A and B. The drug used was a dNTP analog. Other mutants associated to drug and substrate recognition were also mapped by Gibbs *et al.* (1988) in motifs A and B. This is in agreement with a mutant role of these motifs in dNTP binding for the pol α type, as in the case of the pol I type.

Finally, for no-gap DNA polymerases type (pol β), a chemical affinity labeled [^{32}P]8-azido-dGTP peptide on the terminal transferase allowed the mapping of the dGTP binding site; this peptide contained the sequence GHDVD that is part of motif C (Evans *et al.*, 1986) and that resembles both YGDTD and VSRE. This supports the alignment given in Figure 3, even though motif B could not be found with certainty in this pol β type of DNA polymerases. Furthermore, the number of residues between motifs A and C in the terminal transferase DNA-synthesis with the corresponding number in the Klenow pol I is amongst almost 130 residues.

Structural implications of the distance variability between motifs. Figure 3 shows the number of residues contained between each of the motifs for all the DNA pol α sequences and for the Klenow polymerase I fragment. Between motifs A and B a comparable number of residues is found for both pol I and α types, which, between B and C, the pol α sequences generally contain considerably fewer amino acids than their putative Klenow counterparts. This region of the structure encompasses helix F, strands I2 and I3, and helix Q. Apart from the beginning of helix Q, these regions are unlikely to be in contact with the DNA, according to model building studies (Warwicker *et al.*, 1985; Figure 4), suggesting that this region could be consistently

Fig. 4. Alignment of motifs B across nucleotide DNA-dependent DNA polymerases and RNA-dependent DNA polymerases from Human simplex virus (Herpes), vaccinia virus, adenovirus, bacteriophage T7, yeast mitochondria, and retroviruses. Conserved residues are shown in bold. The numbering of the residues is that of the Klenow fragment. Secondary structure elements are indicated above the sequences. Abbreviations: HSV1, Herpes simplex virus type 1; VACCV, vaccinia virus; ADE2, adenovirus type 2; T7, bacteriophage T7; YEAST, yeast mitochondria; HIV1, human immunodeficiency virus type 1; MMTV, murine mammary tumor virus; RSV, Rous sarcoma virus; AMV, avian myeloblastosis virus.

ordered without apparent catalytic basis. The shortest segment between motifs B and C is 29 amino acids in the human pol α sequences; a peptide of this size would easily connect Ile779 at the C terminus of the Klenow motif B to Arg815 at the N terminus of motif C, i.e. connect the loop following beta O with strands 12 and 13 of motif C. In fact, given the 4Å distance between the C_α atoms of Ile779 and Arg815 in the T7 pol ternary structure (Ollis et al., 1985b), 11 residues in helical conformation (strain C_α distance of 1.5 Å) and eight in a coil structure (strain C_α distance of 2.9 Å) could span the segment of 4 Å. This helical portion would favor the maintenance of the N-terminal region of motif C, whose sequence is also numerically conserved in the six type 1 DNA polymerases sequenced (see Figure 2).

The distance between motifs A and B can also vary from 38 to 120 residues in the pol α family; this region corresponds to between 48 and 84 of the Klenow structure. In the pol I family, this distance ranges from 20 to 40 residues. An inspection of Figure 4 suggests the possibility of insertions in this region.

DNA-dependent RNA polymerases

In the family, three conserved regions that matched the three motifs A, B and C could be found. Insertion residues are maintained and the residue lengths between them compare well with those of DNA-dependent DNA polymerases. However, since this family is composed of four members with very two sufficiently distant ones, more sequences are needed in order to narrow the number of conserved regions that can truly functionally connected.

RNA-dependent polymerases

Of the four motifs of viral RNA-dependent RNA polymerases, as well as reverse transcriptase sequences recently identified by Poch et al. (1989), two match motifs A and C of DNA-dependent polymerases. These motifs are the only sequence features shared by all these sequences. The 31% identity alignment of the hepatitis δ reverse transcriptase from Wondrack (Gerbert et al., 1992) with the Herpes simplex virus DNA pol α (Tabor et al., 1987) strongly supports the relationships between DNA- and RNA-dependent polymerases. The existence of these two sequences is also sensible on evolutionary grounds, because reverse transcriptases have recently been postulated to be an essential late evolutary step in going from an RNA-dependent RNA polymerase to a DNA-dependent DNA polymerase (Lazcano et al., 1988). Furthermore, it should be pointed out the hepatitis δ viruses are the only viruses encoding reverse-transcriptase activity that have a DNA genome.

A site-directed mutagenesis experiment has pointed to the catalytic importance of the Asp—Asp doublet in motif C of DNA-dependent polymerases (Bernad et al., 1987). In addition, site-specific mutagenesis of the herpes simplex-DNA polymerase virus reverse transcriptase revealed that mutation of the first conserved Asp of motif C into Asn completely destroyed activity (Larder et al., 1987b). The mutation of the other strictly conserved Asp residue in motif A into Glu had a similar effect.

The sequence length between motifs A and C in the RNA-dependent polymerases averages ~ 90 residues, which is considerably shorter than the 240—320 amino acids in pol Ia, though one sequence from influenza virus contains 718 spacer residues, comparable with some pol Ia (Figure 3). Once again, both N and part of the beta C and strands 12 and 13 are possible candidates for deletion. The absence of all the distances O8 residues suggests that structure can occur in the M and N helices and in the loop following helix O, as they flatly.

Perspectives and conclusion

Argos (1988) previously discussed the conservation of motif C in pol α and RNA-dependent polymerases; we extend this observation to the pol Ia, pol Ib and weak DNA-dependent RNA polymerases. Apparently the fully conserved first Asp of the Asp—Asp doublet of motif C is catalytically more important than the second. We also contend that there is at least another motif upstream (motif A), with another strictly conserved acidic residue in a turn between a beta strand and an alpha helix. In spectral proximity of motif C, genetic experiments suggest that this motif might also be involved in the catalytic process. One of the many possible functional roles for these residues is that the strictly conserved Asp residue in motif C may co-operate with the one of motif A to bind a magnesium ion that would be part of the dNTP binding site. This site would be located at the bottom of a cleft containing the DNA. This cleft, similar to the one of the Klenow fragment, is already apparent in a 4 Å map of T7 DNA-dependent RNA polymerase (Bom et al., 1989). If this is true, the diffences in rates of polymerases could be the result of divergence from a common ancestor, with essentially the same folding, as the consequences of conserved residues in the same spatial arrangement in the complex site could be due to convergent evolution, as observed in nine subtilisin and chymotrypsin active site.

In spite of wide apparent sequence variability (likely to reflect a very ancient divergence), it is very tempting to adopt the strong, structurally unifying principle stating that many polymerases may fold like the known tertiary architecture of the E.coli pol I. Consistent with this hypothesis is the experimental observation that it is possible to change the template or substrate specificity of certain polymerases (type I Mg^{2+} is replaced by Mn^{2+} (see Loeb et al., 1988). This model, however, should be viewed as speculative, even though several lines of evidence point in this unifying conclusion. We believe it deserves attention, because site-directed mutagenesis experiments aimed at probing the catalytic site and template specificity should benefit from our hypothesis, which gives a possible structural framework for

future experiments. If reported crystals of a heterodimer of the reverse transcriptase for human immunodeficiency virus prove to be diffraction worthy (Lowe et al., 1988) and if the high resolution structure of T7 RNA polymerase soon becomes available, then the final test for the hypothesis made here should be forthcoming.

References

[illegible]

[illegible]

EXHIBIT B

Docket No.: NEB-166-PUS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: Jack et al. EXAMINER: Hutson
SERIAL NO.: 10/089,027 ART UNIT: 1652
DATE FILED: March 26, 2002
TITLE: Incorporation of Modified Nucleotides By Archaeon DNA Polymerases And Related Methods

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. §1.131

As a below named inventor, I hereby declare that:

1. My name is Dr. William Jack, Research Director for the DNA Enzymes Division at New England Biolabs Inc. My resume is attached.

2. I have been studying the structure and function of DNA polymerases for over 16 years.

3. I was a member of the group of scientists at New England Biolabs that isolated, characterized, and cloned the first hyperthermophilic archaeal DNA polymerase. Our continuing work with archaeon DNA polymerases identified a surprisingly homogeneous set of enzymes. We claimed this group of DNA polymerases in US Patent 5,500,363. In this patent, the United States Patent and Trademark Office recognized the validity of our claim to a class of archaeon DNA polymerases defined by the DNA encoding the enzyme and its

ability to hybridize under defined conditions to various specified DNA sequences. The group was exemplified by T.litoralis (Vent), GBD (Deep Vent), and 9°N DNA Polymerases.

4. We also found that this group of polymerases had a high degree of amino acid sequence identity. A comparative three-dimensional alignment of members of this group of enzymes showed a high degree of structural conservation, consistent with the observed high degree of primary amino acid sequence identity/similarity. See for example, Vent (Rodriguez, et al., 2000), Tgo (Hopfner, et al., 1999), D. Tok (Zhao, et al., 1999), and KOD (Hashimoto, et al., 2001) DNA Polymerases.

5. The structural equivalence of this group of polymerases is further supported by experiments reported in Example 10 of the above application in which we show that mutation of an analogous residue in Vent and 9°N DNA Polymerases yields enzymes with equivalent acyclonucleotide incorporation efficiencies.

6. We discovered that this group of enzymes is capable of efficiently utilizing acyclonucleotides as substrates. We demonstrated this property using four examples of polymerases within this tightly defined group. Any molecular biologist of ordinary skill in the art would expect from these findings that this property would occur in all members of the enzyme group defined above.

7. Additionally, my colleagues and I have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides, and also for the enhanced incorporation with Vent A488L and 9°N A485L DNA Polymerase mutants. See Gardner, et al. (2004) on page 11841, column 1, paragraph 2 and page 11841, column 2, paragraph

1, respectively.

8. I assert that the combination of the high degree of homogeneity in DNA and amino acid sequences of archaeon DNA polymerases, plus the structural evidence that modification of specific amino acids alters enzyme specificity, would be sufficient to assure a person of ordinary skill in the art that the class of polymerases as defined above will interact with acyclonucleotide substrates as shown in the above application.

9. To further support the above statements, we have conducted additional experiments to confirm that archeon Family B polymerases with an amino acid sequence identity of greater than 30% can utilize acyclonucleotides as a substrate. This data is attached to the present declaration as appendix 1.

9. I further declare under penalty of perjury pursuant to laws of the United States of America that the foregoing is true and correct and that the Declaration was executed by me on:

William E. Jack

Date: 4 May 2006

References:

A. Gardner, C.M. Joyce, W.E. Jack (2004) J. Biol. Chem. 279, 11834-11842.

H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai (2001) J. Mol. Biol. 306, 469-477.

K.P. Hopfner, A. Eichinger, R.A. Engh, F. Laue, W. Ankenbauer, R. Huber, B. Angerer (1999) Proc. Natl. Acad. Sci. USA 96, 3600-3605.

A.C. Rodriguez, H.W. Park, D. Mao, L.S. Beese (2000) J. Mol. Biol. 299, 447-462.

Y. Zhao, D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, J. Kuriyan (1999) Structure 7, 1189-1199.

William Erle Jack
New England Biolabs
240 County Road
Ipswich, MA 01938
(978) 380-7257
(978) 921-1350 (fax)
email: jack@neb.com

RESEARCH INTERESTS

Enzymatic and structural aspects of protein-nucleic acid interactions. Thermostable DNA polymerase kinetics and function.

RESEARCH EXPERIENCE

New England Biolabs (Beverly, MA).

2005-present — Division Head, DNA Enzymes

1987-present — Senior Staff Scientist

Research: Kinetic characterization of thermostable DNA polymerases. Creation and characterization of DNA polymerase variants with altered substrate recognition. Over-expression and characterization of restriction and modification enzymes.

2000-present — New England Biolabs Institutional Biosafety Committee Chair

Rockefeller University (NY, NY) Laboratory of Biochemistry and Molecular Biology.

1983-1987 — Postdoctoral Fellow in the laboratory of R.G. Roeder.

Research: Structural and functional characterization of wild type and mutant forms of *Xenopus* RNA polymerase III transcription factor A. Glucocorticoid hormone-induced transcription enhancement *in vitro*.

Duke University (Durham, NC) Department of Biochemistry.

1977-1983 — Graduate Student in the laboratory of P. Modrich.

Research: Kinetics and thermodynamics of DNA site location, recognition and cleavage by *Eco*RI endonuclease.

EDUCATION

Doctor of Philosophy (Biochemistry), Duke University, 1983 (Paul Modrich, advisor).
Bachelor of Arts (Chemistry), *Magna Cum Laude*, University of Utah, 1977.

TRAINING

2000 — Sixth National Symposium on Biosafety: Prudent Practices for the New Millennium (Conducted by the Centers for Disease Control and Prevention)

PUBLICATIONS

"Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *J. Biol. Chem.* 279, 11834-11842 (2004).

"Acyclic and Dideoxy Terminator Preferences Denote Divergent Sugar Recognition by Archaeon and Taq DNA Polymerases," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.*, 30, 605-613 (2002).

"The Kinetic Mechanism of *Eco*RI Endonuclease," David J. Wright, **William E. Jack** and Paul Modrich, *J. Biol. Chem.* 274, 31896-31902 (1999).

"Determinants of Nucleotide Sugar Recognition in an Archaeon DNA Polymerase," Andrew F. Gardner and William E. Jack, *Nucleic Acids Res.* 27, 2545-2553 (1999)

"Photochemically Initiated Protein Splicing", Sandra N. Cook, **William E. Jack**, Xiaofeng Xiong, Lora E. Danley, Jonathan A. Ellman, Peter G. Shultz, and Christopher J. Noren, *Angew. Chem. Int. Ed. Engl.* **34**, 1629-1630 (1995).

"Three-dimensional Structure of the Adenine-specific DNA Methyltransferase M.TaqI in Complex with the Cofactor S-adenosylmethionine", J. Labahn, J. Granzin, G. Schluckebier, D.P. Robinson, **W.E. Jack**, I. Schildkraut, and W. Saenger, *Proc. Natl. Acad. Sci. USA* **91**, 10957-10961 (1994).

"Expression, Purification, and Crystallization of Restriction Endonuclease PvuII With DNA Containing Its Recognition Site", K. Balendiran, Joseph Bonventre, Roger Knott, **William Jack**, Jack Benner, Ira Schildkraut, and John E. Anderson, *Proteins: Structure, Function, and Genetics* **19**, 77-79 (1994)

"Protein Splicing Elements: Inteins and Exteins--A Definition of Terms and Recommended Nomenclature", Francine B. Perler, Elaine O. Davis, Gary E. Dean, Frederick S. Gimble, **William E. Jack**, Norena Neff, Christopher J. Noren, Jeremy Thorner and Marlene Belfort, *Nucleic Acids Res.* **22**, 1125-1127 (1994).

"Characterization of a DNA Polymerase from the Hyperthermophile Archaea *Thermococcus litoralis*", Huimin Kong, Rebecca B. Kucera and **William E. Jack**, *J. Biol. Chem.*, **268**, 1965-1975 (1993)

"Protein Splicing Removes Intervening Sequences in an Archaea DNA Polymerase", Robert A. Hodges, Francine B. Perler, Christopher J. Noren and **William E. Jack**, *Nucleic Acids Res.* **20**, 6153-6157 (1992).

"Intervening Sequences in an Archaea DNA Polymerase Gene", Francine B. Perler, Donald G. Comb, **William E. Jack**, Laurie S. Moran, Boqin Qiang, Rebecca B. Kucera, Jack Benner, Burton E. Slatko, Donald O. Nwankwo, S. Kay Hempstead, Clotilde K.S. Carlow and Holger Jannasch, *Proc. Natl. Acad. Sci. USA* **89**, 5577-5581 (1992).

"Overexpression, Purification and Crystallization of *Bsu*RI Endonuclease", **William E. Jack**, Lucia Greenough, Lydia F. Dorner, Sloung-yong Xu, Teresa Szczelkun, Aneel K. Aggarwal and Ira Schildkraut, *Nucl. Acids Res.* **19**, 1825-1829 (1991).

"Nucleotide Sequence of the *EcoRI* Restriction-Modification System: Separate Strand-Specificity Domains in the Methyltransferase", Mary C. Looney, Laurie S. Morse, **William E. Jack**, George R. Feehery, Jack S. Benner, Barton E. Slatko and Geoffrey G. Wilson, Gene **80**, 193-208 (1989).

"M.EcoRI methylates adenine in both strands of its asymmetric recognition sequence", David Landry, Mary C. Looney, George R. Feehery, Barton E. Slatko, **William E. Jack**, Ira Schildkraut, and Geoffrey G. Wilson, *Gene* **77**, 1-10.

"Mechanism of specific site location and DNA cleavage by *EcoRI* endonuclease", Brian J Terry, **William E. Jack** and Paul Modrich, *Gene Amplif Anal.* volume 5, pp. 105-118 (1987)

"Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis", Brian J. Terry, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **260**: 13130-13137 (1985).

Participation of Outside DNA sequences in the EcoRI Endonuclease Reaction Pathway, **William E. Jack**, dissertation (Duke University, Durham, NC) (1983).

"Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences", Brian J. Terry, **William E. Jack**, Robert A. Rubin and Paul Modrich, *J Biol. Chem.* , **258**, 9820-9825 (1983).

"Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence", **William E. Jack**, Brian J Terry, and Paul Modrich, *Proc Natl. Acad Sci U S A* **79**: 4010-4014 (1982)

"DNA determinants important in sequence recognition by *EcoRI* endonuclease", A-Lien Lu, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **256**: 13200-13206 (1981).

"Structures and Mechanisms of *EcoRI* DNA Restriction and Modification Enzymes", **William E. Jack**, Robert A. Rubin, Andrea Newman and Paul Modrich, *Gene Amplif. Anal.* volume 1, p 165-179 (1981).

US PATENTS

"Use of site-specific nicking endonucleases to create single-stranded regions and applications thereof," **William E. Jack**, Ira Schildkraut, Julie Forney Mertes, US Patent 6,660,475, December 9, 2003.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,834,285, November 10, 1998.

"Modified protein comprising controllable intervening protein sequences or their elements methods of producing same and methods for purification of a target protein comprised by a modified protein," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, Christopher J. Noren, Shaorong S.C. Cheng, Eric Adam, Maurice Southworth, US Patent 5,834,247, November 10, 1998.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,500,363, March 19, 1996.

"Modification of protein by use of a controllable intervening protein sequence," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, US Patent 5,496,714, March 5, 1996.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,352,778, October 4, 1994.

"Purified thermostable DNA polymerase obtainable from thermococcus litoralis," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,322,785, June 21, 1994

doi:10.1016/j.jmb.2006.07.028 Available online at http://www.sciencedirect.com on BrainJ. Mol. Biol. (2006) 360, 447–462

JMB

Crystal Structure of a Pol α Family DNA Polymerase from the Hyperthermophilic Archaeon *Thermococcus sp.* 9°N-7

A. Chapin Rodriguez†, Hee-Won Park†, Chen Mao and Lorena S. Beese*

Department of Biochemistry Duke University Medical Center, Durham NC 27710, USA

The 2.25 Å resolution crystal structure of a pol α family (family B) DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol) provides new insight into the mechanism of pol α family polymerases that include essentially all of the eukaryotic replicative and viral DNA polymerases. The structure is folded into NH_2-terminal, editing 3′-5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures, bacteriophage RB69 and the recently determined *Thermococcus* gorgonarius, but differs in their relative orientations and conformations.

The 9°N-7 polymerase domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of free dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparisons of the orientations of the fingers and exonuclease (exo)domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the DH-terminal domain of 9°N-7 pol may be structurally related to an RNA-binding motif, which appears to be conserved among archaeal polymerases. The presence of such a putative RNA-binding domain suggests a mechanism for the observed interrogation of bacteriophage T4 DNA polymerase synthesis by binding to its own mRNA. Furthermore, conservation of this domain could indicate that such regulation of pol expression may be a characteristic of archaea. Comparison of the 9°N-7 pol structure to its mesostable homolog from bacteriophage RB69 suggests that thermostability is achieved by shortening loops, forming two disulfide bridges, and increasing electrostatic interactions at subdomain interfaces.

Keywords: Archaea; X-ray structure; replication; exonuclease; family B DNA polymerase

*Corresponding author

Introduction

DNA polymerases catalyze the template-directed addition of nucleotides onto the 3′-OH group of the DNA primer terminus. These enzymes replicate DNA with the required accuracy essential for genomic

†Contributed equally to the manuscript.

Abbreviations used: pol, polymerase; T_m, *Thermococcus gorgonarius*; ddNTP, dideoxyribonucleotide.

E-mail address of the corresponding author: sbeese@structure.duke.edu

0022-2836/$30/$see front matter

mic stability, that generate sufficient mutations to stimulate and maintain evolution. Unlike Eucarya and Bacteria, relatively little is known about DNA replication in Archaea (Poole et al., 1995), one of the three major evolutionary lineages of life (Woese et al., 1990). Archaea play a significant role in the biosphere, accounting for up to 30% of the biomass in certain Antarctic waters (De Long et al., 1994), and exhibit much greater diversity than had originally been suspected (Burns et al., 1996). Many characterized archaeal species are adapted to live in environments of extreme temperature, pressure, salinity, and/or pH such as hydrothermal vents, and hot springs (Ross & Adams, 1995).

Although archaeal cells share many morphological features with Eucarya, archaeal proteins involved in gene expression including DNA replication, transcription, and translation have been found to be similar to those from Eucarya (Edgell & Doolittle, 1997; Puhl et al., 1996) in particular most of the archaeal DNA polymerases that have been sequenced belong to the α-like polymerase family (family B) that includes essentially all the eukaryotic replicative and viral DNA pols (Braithwaite & Ito, 1993; Edgell et al., 1997).

Crystal structures exist for DNA pols from each of four families pol I (family A), pol α (family B), pol β (family X) and reverse transcriptase interviewed by Joyce & Steitz, 1994; Doublié et al., 1999). Although pols from different families are structurally quite diverse, several common features have emerged. The pol domains from each resemble a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of *Escherichia coli* pol I (Klenow fragment (Ollis et al., 1985). All polymerases appear to share the same mechanism for nucleotidyl transfer involving two divalent metal ions (reviewed by Beese & Steitz, 1994). In addition, based on structures containing DNA and dNTP bound to pols from pol I, pol β, and reverse transcriptase inhibitors, a conformational change in the fingers subdomain from an open to a closed conformation is proposed to occur during the catalytic cycle (reviewed by Doublié et al., 1999).

The pol α family polymerases are of medical importance as targets for development of antiviral and antitumor therapeutics. For example, human pol α is a target in the treatment of acute myelogenous leukemia and chronic lymphocytic leukemia (Keating et al., 1982; Robertson & Plunkett, 1985) and a variety of nucleotide analogs with antitumor activity inhibit strand elongation by pol α (Huang & Plunkett, 1991; Cazaubon & Plunkett, 1995). Furthermore, polymerases, particularly those that are thermostable, have a number of critical biotechnological applications ranging from PCR to cloning and DNA sequencing. Despite their biological, medical and biotechnological importance, the pol α class of polymerases has not been structurally as well characterized as other DNA polymerase families.

Here we report the 2.25 Å resolution crystal structure of a pol α family DNA polymerase from the hyperthermophilic marine archaeon Thermococcus sp. 9°N-7 (9°N-7 pol). Thermococcus sp. 9°N-7 was isolated from a 9-deep-sea vent at 9° N latitude off the East Pacific Rise (Southworth et al., 1996). The structure is folded into NH2-terminal, editing 3′-5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures Bacteriophage RB69 (Wang et al., 1997) and the recently determined *Thermococcus gorgonarius* (Tgo) (Hopfner et al., 1999), but differs in their relative orientation and conformation.

The pol domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the two other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparison of the architectures of the fingers and exonuclease domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can serve as a unit, and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the NH_2-terminal domain of 9°N-7 pol is structurally homologous to the RadA/Sms RNA-binding motif with an exposed patch of aromatic amino acid residues. Bacteriophage T4 DNA pol, which is homologous to 9°N-7 pol, is known to bind its own mRNA and repress its own synthesis. The homology relationships of the RNA-binding motif suggest a structural basis for this regulatory mechanism. Furthermore, the conservation of this domain in other archaeal pols suggests that such autogenous regulation of pol expression may be general for archaea.

Results and Discussion

Crystal structure of *Thermococcus* sp. 9°N-7 pol

The structure of the full-length, 775-residue enzyme showing the double mutations D141A and D144A(6x) was determined using the multiple isomorphous replacement method to a resolution of 2.25 Å. The current model has an R factor of 23.9% (R_{free} = 26.8%) (Table 2). A Ramachandran plot of the model shows 86.5% of the residues in the most favored region and the remainder in additional allowed regions (12.4%) and generously allowed regions (0.8%). A total of 37 residues are not traced in the model and lie in regions of poorly defined electron density. The first of these gaps

Table 1. Crystallographic data collection and refinement statistics

Data	Native[a]	Native	Native	Hg[a]	Hg	Hg	Hg	Hg	Hg	Hg
A. Data collection and phasing[b]										
Peak	1.0	0.98	0.9	1.0	1.0	0.9	0.9	0.9	0.9	0.9
Inflection	1.0	0.98	0.9	1.0	1.0	0.9	0.9	0.9	0.9	0.9
Resolution (Å)	20–2.4	20–2.1	20–2.1	20–2.5	20–2.5	20–2.5	20–2.5	20–2.5	20–2.5	20–2.5
Unique reflections	19,648	31,644	31,631	16,218	16,218	16,218	16,218	16,218	16,218	16,218
Completeness[c] (%)	97.0 (91.1)	97.5 (84.6)	97.5 (84.6)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)
Redundancy[c]	9.9 (9.1)	10.2 (10.0)	10.2 (10.0)	10.2 (10.0)	10.2 (10.0)	10.2 (10.0)	10.2 (10.0)	10.2 (10.0)	10.2 (10.0)	10.2 (10.0)
R_{sym}[c,d]	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
I/σ[c]	17.7	17.7	17.7	17.7	17.7	17.7	17.7	17.7	17.7	17.7
R_{iso}										
Phasing power[e]				1.1	1.1	1.1	1.1	1.1	1.1	1.1
B. Refinement (Native)										
Resolution limits (Å)	20–2.4	20–2.1								
Reflections in working set	18,635	30,062								
Reflections in test set (5)	979	1,582								
Cutoffs	$F \geq 0$	$F \geq 0$								
R_{cryst}[f] (%)	21.8	20.4								
R_{free}[g] (%)	27.3	25.4								
Protein atoms	2,436	2,436								
Water molecules	154	154								
Mean B ($Å^2$)	29.4	29.4								
rms deviations										
Bonds (Å)	0.006	0.006								
Angles (°)	1.3	1.3								

[a] Quantities in bold indicate the highest resolution shell.
[b] Values in parentheses are for the highest resolution shell.
[c] $R_{sym} = \sum|I - \langle I \rangle| / \sum I$, where I refers to the intensity of an individual measurement and $\langle I \rangle$ to the mean intensity for all symmetry-related observations.
[d] I/σ is the mean intensity divided by its estimated error based on counting statistics.
[e] Phasing power = $\sum|F_{H(calc)}| / \sum|F_{obs} - F_{H(calc)}|$, where $F_{H(calc)}$ is the calculated heavy-atom structure factor and F_{obs} is the observed structure factor, respectively.
[f] $R_{cryst} = \sum||F_{obs}| - |F_{calc}|| / \sum|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.
[g] R_{free} is equivalent to R_{cryst}, but for the 5% of reflections excluded from refinement and not used in the working set.

cterus at the bottom of the palm domain (residues 568–575), and the remainder are within the thumb regions that are frequently observed to be partially disordered in apo polymerase structures, as is also the case here (e.g. Ollis et al., 1995; Kiefer et al., 1998). Although no disulfide bridges were included in the refinement, four Cys residues showed anomalous peaks in a difference Fourier map and side-chain distances and angles consistent with two disulfide bridges (Cys408-Cys411, Cys508-Cys511).

The structure of 9°N-7 pol reveals features common to all DNA pol structures as well as those that may be unique to archaeal pols. The overall shape of the enzyme can be described as a claw with a central hole that is lobed into NH$_2$-terminal, 3′-5′ exonuclease, and polymerase domains (Figure 1(a) and (b)). Like all other pols of known structure, the pol domain resembles a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of E. coli pol I (Klenow fragment) (Ollis et al., 1985). 9°N-7 pol is similar in structure to the pol α family polymerase from the mesophilic bacteriophage RB69 (RB69 pol) (Wang et al., 1997), although a number of those subdomains are shorter than in RB69 pol (Figure 1(c)). Nearly all those sequence length differences are attributable to loop segments that are fewer and shorter in the hyperthermostable 9°N-7. As was first observed in the RB69 pol structure (Wang et al., 1997), the 3′-5′ exonuclease domain lies on the opposite side of the palm in comparison to pol I family polymerases. This domain arrangement is also seen in 9°N-7 pol and in Taq pol (Kiefer et al., 1998), indicating that this motif is likely to be general for the pol α family. The structural similarity between 9°N-7 and RB69 pols is significant given the low sequence identity (<20%) in all but the active-site (palm) region, where sequence identity is 42% (Figure 2). Similar results hold for sequence alignments between 9°N-7 and human pol δ.

NH$_2$-terminal domain

Many of the members of the pol α polymerase family, including archaeal pols, bacteriophage T4 and RB69 DNA pols, have an NH$_2$-terminal domain that is not observed in the pol I family. T4 pol is known to control its synthesis in vivo by a mechanism of autogenous regulation (Tinker et al., 1990). The mRNA-binding activity has been located to within the first 105 residues of the pol (Wang et al., 1996), but the structure of a fragment comprising residues 1–288 of T4 pol failed to suggest a structural basis for RNA binding (Wang et al., 1996). Here, we note that certain structural similarities between the homologous regions in the 9°N-7 pol and the T4A RNA-binding protein may provide a reference for RNA binding by T4 pol.

The NH$_2$-terminal domain of 9°N-7 pol can be considered in three modules based on connectivity of folding (Figure 3(a)). The first module comprises residues 1–51, a three-stranded β-sheet that interacts extensively with the 3′-5′ exonuclease domain via predominantly electrostatic interactions. Residues 52–56 act as a flexible linker connecting the first module to the second (residues 57–128). The third module comprises residues 308–376.

The second module is folded into a βαββαβ motif, with two short β-strands, 5 and 6, inserted between the second and third elements. This motif occurs in a variety of proteins, and forms the basis for the most prevalent RNA binding motif, the RNA recognition motif (RRM). The RRM is present in the RNA-binding domains of hnRNP-A1, spliceosomal protein U1A and U2B″, and the sex lethal protein (Burd & Dreyfuss, 1994). Although an alignment of the NH$_2$-terminal domains of archaeal pols (Figure 3(b)), together with T4 and RB69 pols, shows that they lack the RNP1 and RNP2 sequence motifs that characterize the RRM (Burd & Dreyfuss, 1994), a number of highly conserved and consistent residue substitutions emerges. Most of these residues fall in a cluster on the surface of the NH$_2$-terminal domains of 9°N-7 and RB69 pols which therefore could mark the location of an RNA binding site atop the β-sheet platform on the face away from helix A (Figure 3(c)).

Both a sequence alignment (Figure 3(b)) and a structural comparison (Figure 3(c)) reveal that T4 and RB69 pols lack helix A and strand 2 of the βαββαβ motif, perhaps explaining why no suggestive structural homologues to RNA-binding folds could be identified (Wang et al., 1996, 1997).

Experiments are needed to determine whether the NH$_2$-terminal domain of 9°N-7 pol binds RNA. Although the βαββαβ motif occurs in proteins that are not thought to interact with RNA (Burd & Dreyfuss, 1994), we find its presence in the NH$_2$-terminal domain of 9°N-7 pol, in a region known to bind RNA in T4 pol (Wang et al., 1996), to be highly suggestive of this. RNA binding capability could hold for other archaeal pols as well, since sequence alignment of NH$_2$-terminal domains (Figure 3(b)) suggests that they share the βαββαβ motif.

We further speculate that pols as T4 pol binds its mRNA to down-regulate its own synthesis, such autogenous regulation of pol expression might occur in archaea. Autogenous gene regulation is well documented in bacteria, and has at least one precedent in archaea. It has been identified in the synthesis of the RNA1 ribosomal protein in *Methanococcus vannielii* (Hasenbank et al., 1994), and postulated for a ribosomal gene cluster from the halophile *Halobacterium salinarium* (Shimmin & Dennis, 1990). It is interesting that there is no structural evidence that such regulation extends to eukaryotes, as human pol α shows no significant sequence homology to the NH$_2$-terminal sequences aligned in Figure 3(b).

3′-5′ Exonuclease domain

This domain is responsible for binding single-stranded DNA and excising mismatched bases in the elongated primer strand. The structure

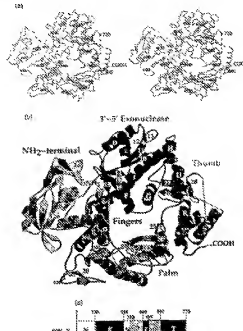


Figure 1. Structure of the Thermococcus sp. 9°N-7 DNA polymerase. The NH2-terminal and 3'-5' exonuclease domains are colored yellow and green, respectively. The polymerase domain is divided into palm (brown), thumb (red), and fingers (blue) subdomains. Turns highly conserved carboxylate groups (Dx01, Lx040, Yx041) mark the polymerase active site. (c) Homologue of the C5' ends. Every fifth C' is numbered. Starters from fourteen distinctive regions of the proteins. (b) Ribbon diagram with secondary structure elements defined according to RB69 (Kohlstaedt & Steitz, 1992). Fifty essential domains: 1, 1-30; 2, 10-22; 3, 33-35; 4, 37-49; A, 60-63; 5, 53-58; 6, 61-64; 7, 67-76; 8, 78-85; 9, 90-101; B, 95-121; 9, 109-113; C, 116-125; 1, 142-144; E, 204-206; 2-7 noncontiguous domains. 10, 135-146; 11, 159-162; 12, 168-172; 13, 180-183; D, 187-201; 14, 209-268; E, 218-225; 15, 240-244; 16, 247-251; 17, 256-259; 9, 260-264; G, 273-285; 18, 293-305; L, 308-337; Polymerase domains: I, 374-376; 19, 397-404; M, 408-415; 19, 435-443; 20, 460-462; N, 468-469; O, 475-496; P, 507-522; 21, 535-539; Q, 543-547; Q, 585-587; 22, 595-600; 23, 603-606; 24, 617-635; S, 636-651; 1649-665 noncontiguous; T, 685-690; glutaredoxin 26, 692-695; U, 677-683; 27, 698-703; 28, 714-716; V, 758-764; W, 782-784; (c) homologous corresponding the thioredoxins of Thermococcus sp. 9°N-7 Total bacteriophage RB69 DNA polymerase. The domain boundaries for 9°N-7 pol were determined based upon a structure-based sequence alignment with RB69 pol (Figure 7) as defined for the RB69 pol (Wang et al., 1997).

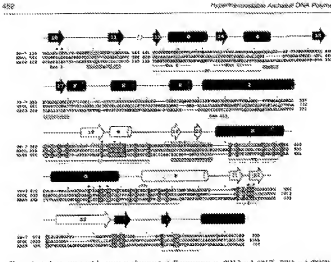


Figure 2. A three-way partial sequence alignment of *Thermococcus* sp. 9°N-7 pol (9N7), RB69 pol (RB69), and human pol α (HUM). Dashes indicate gaps in the alignment, and segments not aligned are represented as twelve residue spans within brackets. Ticks mark every 50 species. The 9°N-7 and RB69 pol alignment is based upon the crystal structures. The 9N7, and RB69 segments is from Wang et al. (1997), except for a few short segments realigned based upon the three sequences shown here. Indicated below the sequences, and boxed in yellow are conserved residues in the exonuclease (Beese et al., 1993) and polymerase (Wang et al., 1997) domains. The secondary structure elements in 9°N-7 pol, as defined by DSSP, are given above the sequences. The structural elements are labeled according to the scheme described in regard to figure 1. Shown in purple is the 9°N-7 pol segment are the archaeal polymerase motifs described by Edgell et al. (1997). Residues within the polymerase domain that are associated in the three sequences blue boxes; residues discovered in the section on dNTP binding, blue asterisks. The two disulfide bridges in the palm (CA38-C41, C74-C477) are shown tube-satellite.

reported here is that of a mutant of 9°N-7 pol lacking detectable exonuclease activity which was engineered to prevent degradation of DNA substrates during subsequent co-crystallization experiments. This 9°N-7exo⁻ pol was obtained by making two point mutations, (D141A, E143A) in the first 3'-5' exonuclease domain of many DNA pols (Derbyshire et al., 1986; Bernad et al., 1992). In the Klenow fragment (KF) of *E. coli* DNA pol I, these residues (D368, E357) are responsible for binding the catalytic metals and for hydrogen bonding with the 3'-OH of the terminal nucleotide residue of the substrate DNA (Beese & Steitz, 1995).

Aside from loop segments that are shorter than those observed in RB69 pol (see below), the topology of the exonuclease domain in 9°N-7 pol is very similar to that of RB69 pol. The domains superimpose in the central β-sheet, containing the active site, with a root mean square deviation (rmsd) of 0.98 Å (85 Cᵅ atoms). The metal binding residues not mutated in 9°N-7exo⁻ pol, D215 and D218, superimpose almost exactly on the corresponding RB69 pol residues (D222, D207).

It is now possible to assign a structural context to the four archaeal sequence motifs identified by Edgell *et al.* (1997). Three of the regions (A-C) lie within the exonuclease domain (Figure 3). Motif A forms part of the central β-sheet containing the active site; B, part of a solvent-exposed loop; and C, part of a five-stranded β-sheet nearly perpendicular to the central β-sheet. The fourth motif resides in the palm (see below).

Pol domain

This domain is responsible for the template-directed polymerization of dNTPs onto the growing primer strand of duplex DNA. Like other polymerases of known structure, the pol domain can be further divided into palm, fingers, and thumb subdomains. While the structure of the fourth of 9°N-7 and RB69 pols are highly similar, differences arise in the palm and fingers. Some of these differ-

sites correspond to features that appear unique to archaeal pols, while others support a hypothesis that a conformational change occurs in the fingers as part of the catalytic cycle.

Palm subdomain

The palm, which contains the active site for polymerization, shows a high degree of structural similarity to the palm subdomains of other DNA polymerases. It is as structurally similar to pol I family polymerases as to those of the pol α family. Its one distinction from RB69 pol around the active site (blue region in Figure 4[a]) is 6.36 Å Cβ–Cβ atoms. Together with the Tgo pol structure (Hopfner et al., 1999), this structure confirms for archaea the conservation of a common catalytic core. A significant difference between the palm subdomains in 9°N-7 and RB69 pols are the two disulfide bridges present in 9°N-7 pol, one joining Cys285 and 442 and another joining Cys504 and 509 (Figure 4[b]). Both the shortened loops and at least one disulfide bridge appear common to archaeal pols (see above). Indeed, the region containing one of the Cys residues in a disulfide bridge (C442) corresponds to the highly conserved archaeal motif G (Edgell et al., 1997; Figure 2). The Tgo pol structure shows the corresponding Cys residues to be "poised" for disulfide formation, but still in reduced form.

Until recently it was believed that all pols share a catalytic "triad" of carboxylate residues in the active site in the palm (Delarue et al., 1990). Wang et al. (1997) also recognized that only two of the carboxylate residues are invariant. The invariant carboxylates in 9°N-7 pol are D408 and D540. The third member of the triad, present as D580 in 9°N-7 pol, is not essential: mutations of the corresponding residue (D1082N) in human pol α retain catalytic function (Copeland et al., 1992). D540 in 9°N-7 pol may nevertheless be involved in binding the divalent metals required for catalysis. Mg^{2+} is normally the optimal metal for human pol α activity. The pol α D1082N mutant shows greater catalytic efficiency and fidelity with Mn^{2+} rather than Mg^{2+} (Copeland & Wang, 1993).

D540 in 9°N-7 pol interacts with the hydroxyl group of Y538 that is within hydrogen-bonding distance to D540. Substitution of this residue in Pol III human pol α (Y1262G) causes only minor effects on catalysis but alters the pol α metal affinity akin to the pol α D1082 mutation (Copeland & Wang, 1993). It seems likely that the hydroxyl moiety of Y538 in 9°N-7 pol helps to fix D540 in position for Mg^{2+}-specified binding. Consistent with this function is the strict conservation of Y538 among pol α family members (Braithwaite & Ito, 1993).

Fingers subdomain

The fingers subdomain of 9°N-7 differs in topology and relative conformation from RB69. The fingers of 9°N-7 pol use a simple hairpin-and-helix, as in Tgo pol (Hopfner et al., 1999), whereas in the fingers of RB69 pol, the coil region is expanded with more secondary structure elements (fingers 2 and 3). The shorter fingers of 9°N-7 pol are conserved among the archaeal pols aligned by Edgell et al. (1997). It is possible that the fingers of archaeal pols define a minimal functional unit.

Different positions of the fingers subdomain relative to the palm are observed in the 9°N-7 and RB69 pol structures (Figure 5[a]). The fingers of Tgo pol (Hopfner et al., 1999) show a position intermediate between that in 9°N-7 and RB69 pols, when the palm subdomains of all three enzymes are aligned. It is interesting to note that the fingers subdomain of polymerases in the pol I family adopt different positions during the catalytic cycle (reviewed by Doublié et al., 1998). An open position corresponds to that seen in the apoenzyme form (Ollis et al., 1988; Kim et al., 1995; Kiefer et al., 1998; Kiefer et al., 1997) and the form bound to duplex DNA (Kim et al., 1995; Kiefer et al., 1998). A closed conformation has been observed in the ternary replication complexes of bacteriophage T7 pol (Doublié et al., 1998) and Klentaq (Li et al., 1998) with bound DNA and dNTP. An analogous conformational change has been observed in inverse transcriptase (Huang et al., 1998) and pol β (Pelletier et al., 1994). In the closed conformation the fingers rotate towards the palm to form a binding pocket for dNTPs.

The differences in position of the fingers subdomain in the three pol α family crystal structures suggest that the fingers of pol α family pols move during catalysis, analogous to that observed for the other polymerase families. It is interesting to note that if this is the case, there must be a corresponding movement in the position of the 3′-5′ exonuclease domains not required in the other polymerase families as will be discussed below in the position of the fingers in 9°N-7 pol some clearly aggressive/open a closed conformation. It is not clear why they would adopt a position previously observed only in ternary complexes with bound dNTP and DNA. The fingers of 9°N-7 pol may be stabilized in this conformation because of a salt-bridge between R308 in the palm and R460 on helix O of the fingers. These residues are highly conserved among archaeal pols (Edgell et al., 1997) and both pol I and pols α families (Braithwaite & Ito, 1993). The corresponding salt-bridge does not form in polymerases of the pol I family because the fingers helix O lies too far from the palm. The fingers of Tgo pol, in fact, are rotated slightly away from the active site, relative to 9°N-7 pol, such that the E309–R457 salt bridge cannot form. Another possible explanation for the difference in finger positions are the disulfide bridges present in 9°N-7 pol but absent in the Tgo pol structure and in pol I family structures. At least one of the disulfides (Cys426-442) in 9°N-7 pol could be directly involved in orienting the fingers relative to the palm (Hopfner et al., 1999).

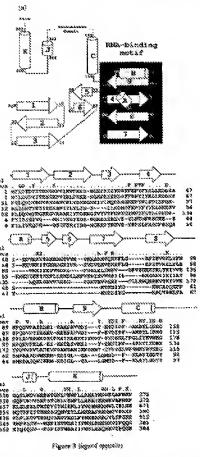


Figure 3 (legend opposite)

Model for DNA and dNTP binding

Based on the high degree of structural homology of the palm subdomain between 9°N-7 and pol I, ternary pols. DNA and dNTP substrates from the bacteriophage T7 pol ternary complex (Doublie et al., 1998) were modeled into the 9°N-7 pol active site. The model shown in Figure 8 provides further

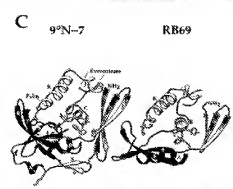


Figure 2. The RNA-binding motif in the NH$_2$-terminal domain of 9°N-7 pol. (a) Topology diagram of the complete NH$_2$-terminal domain (residues 1–428, 9°N-7). The RNA-binding motif (ribbon) located in the DNA polymerase motif (Bork & Doolittle, 1994), as boxed, the sequence alignment of the NH$_2$-terminal domains of 9°N-7 pol, RB69 pol, T4 pol, and archaeal pols. Alignments of 9°N-7 pol and TA/RB69 is based upon the crystal structures, and that of 9°N-7 pol and the other archaeal polymerases is based upon sequence alignment of 13 sequences among twelve considered by Edgell et al. (1997) (data not shown). The aligned polymerase sequences were performed with the PILEUP algorithm at the GCG package University of Wisconsin Genetic Computer Group. Secondary structure elements corresponding to 9°N-7 pol are given above the sequence. A consensus sequence was derived for the archaeal polymerases at those positions where at least 92% of the 13 sequences showed the same residues. Boxed in yellow are those residues conserved between the archaeal consensus and both bacteriophage (T4, RB69) sequences. Position 367 on 9°N-7 pol is starred (see the text for discussion). Abbreviations are as follows. PFU, *Pyrococcus furiosus* DSM3; *Pofu*, *Pyrococcus* sp. GB-D; *Pohor*, *Pyrococcus horikoshii* OT3; *Aepyr*, *Aeropyrum pernix*; *Sulac*, *Sulfolobus acidocaldarius*; *Sulso*, *Sulfolobus solfataricus*; *Metjan*, *Methanococcus jannaschii*; *Pyrae*, *Pyrodictium occultum* B1 (9). Ribbon representation of the NH$_2$-terminal domain of 9°N-7 (left) and RB69 (right) pols. Least-squares C^{α} superposition was performed over the region of 9°N-7 pol including strand 4, part of strand 5, and helices B and C, and the domains were separated for subtlety side comparison. Shown in green is the flexible RNA-binding motif. Charged and aromatic exposed conserved residues are shown with green side-chains, and yellow side-chains correspond to the residues boxed in yellow in (a). The loop between β strands 7 and 8 in 9°N-7 pol corresponds to the conformationally variable loop 3 in the canonical RNP motif (Hunter *et al.*, 1995).

evidence that the position of the fingers in 9°N-7 pol exists closely approximates a closed conformation and their position in RB69 pol approximates an open conformation. This model of a loosely complex for a pol a family polymerase places the dNTP within hydrogen-bonding distance of residues on the fingers O helix that are highly conserved and known by mutagenesis to be functionally important. The corresponding residues on fingers helix P of the RB69 pol are further away and cannot directly interact with dNTP.

The model places residues Y438 and Y494 near the deoxyribose moiety of the incoming dNTP. These residues appear to be functionally analogous to F880 and Y926 of T7 pol, which are responsible for discriminating between deoxy- and ribonucleotides (rNTPs). Y438 is invariant among the pol a family. In the alignment by Braithwaite & Ito (1993) and nearly invariant (one exception) among archaeal pols aligned by Edgell et al. (1997). Mutation of the corresponding residue (Y412) to V41 to an exonuclease-deficient *Thermococcus litoralis* (Vent) pol causes a 320-fold loss of discrimination against rNTPs. The observation may appears to be the functionally important moiety, as mutating Y412 in the consensus wild-type discrimination levels (Gardner & Jack, 1999).

Y506 in T7 pol (F762 in Klenow fragment) has been dubbed the "ribose selectivity sole" (Tabor & Richardson, 1998). A Phe residue at this position confers selectivity against incorporation of dideoxynucleotides (ddNTPs), whereas a Tyr residue in this position allows efficient incorporation of both nucleotide species. The presence of Tyr (Y439) in this position in 9°N-7 pol suggests the ability to incorporate dideoxynucleotides, as do Vent (Gardner & Jack, 1999) and human pol a (Copeland et al., 1992). In fact, Tyr is invariant at this position among the archaeal pols aligned by Edgell et al. (1997), and highly conserved in the pol a family aligned by Braithwaite & Ito (1993).

The model of a ternary complex with dNTP and DNA places residues S489 and K487 in hydrogen-bonding distance from the triphosphate moiety of

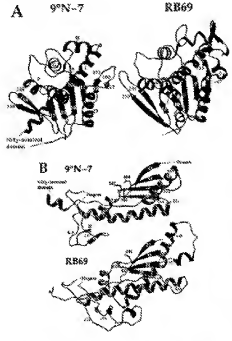


Figure 4. Comparison of 9°N-7 and RB69 pols in different functional states to indicate large segments that are shorter in 9°N-7 pol. Lettering conventions as in Figure 3. Each structure was positioned over the region in blue, and the domains were separated for ease-by-side comparison. Loop regions are shown in magenta and their relative endpoints are labeled. (a) Comparison of the exonuclease domains. Indicated with purple asterisks are the active site carboxylates mutated to Ala in the case of the 9°N-7exo- pol used in the study. (b) Comparison of the palm domains. The three active-site carboxylate groups are depicted with side-chains.

the incoming dNTP. Both of these residues are mutated in the pol α family (Braithwaite & Ito, 1993), and nearly mutated from exonuclease active archaeal pols (Edgell et al., 1997). Mutation of the corresponding residues (Y408, K410 in Vent (exo-) pol severely decreases enzyme activity (Gardner & Jack, 1999).

Concerted domain movement

The difference in position of the fingers subdomains in 9°N-7 and RB69 pols is part of a larger conformational change involving the 3′-5′ exonuclease and NH_2-terminal domains. Comparing these two pol structures shows that in one of the pols, an essentially rigid-body rotation has occurred involving three of the five polymerases. This concerted movement affects both the position of the fingers relative to the pol active site (open versus closed conformations) as well as the position of the exonuclease active site relative to the pol active site. The 9°N-7 and RB69 pol structures map approximate different states along the reaction pathway corresponding to DNA synthesis and 3′-5′ exonucleolytic proofreading activities.

When these two polymerases are aligned in the palm (the blue region in Figure 4b), the exonuclease and fingers are displaced between the proteins (Figure 5a). If the enzymes are aligned in the exonuclease domain (see Figure 4a), the fingers superimpose almost exactly (Figure 5b). Aligning these 2 pairs to an exonuclease-based alignment also brings the first module (residues 1–81) of the NH_2-terminal domains into identical positions (not shown). The joint evolution of the first NH_2-terminal module and the exonuclease may reflect the need to maintain ionic networks at the interface. There are two five-membered ionic net-

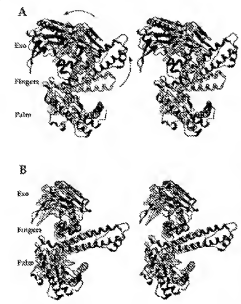


Figure 5. Least-squares C^{α} superpositions of 9°N-7 and RB69 pols in the (A) palm subdomain or the exonuclease domain. The 9°N-7 pol backbone is shown in yellow, and its active-site carboxylate groups in gold. The RB69 pol backbone is shown in green, and its active-site residues in magenta. The catalytic Mg-ions of the exonuclease domains in light blue (9°N-7 pol) or dark blue (RB69 pol) to allow tracking of the domain rotation. The proofer regions used in the palm and exonuclease superpositions are shown in Figure 6. The NH_2-terminal domains have been omitted for clarity. Arrows to (a) indicate the direction of fingers and exonuclease movement when moving from (a) to (b).

worms formed between the first module and exonuclease (Figure 7). In addition, a three-membered network is formed between the third NH_2-module (R586) and the exonuclease (Figure 7). This network is conserved among nearly all archaeal pols (Edgell et al., 1997), but more so present in RB69 pol.

Comparisons of the Tgo pol structure (Hopfner et al., 1999) with that of 9°N-7 and RB69 pols using palm and exonuclease-bound superpositions gives results similar to those in Figure 5, providing further support for the notion of a conserved domain movement.

A model was constructed for the RB69 pol (Wang et al., 1997) showing how substrate DNA could shuttle between the pol and exonuclease active sites. When 9°N-7 and RB69 pols are aligned in the palm, the exonuclease active site in the former is filled out and away from the pol active site, making it compatible for the DNA to shuttle. The exonuclease position in RB69, but not that in 9°N-7 pol, is therefore consistent with an editing conformation. It is interesting that this creative motion also ensures that the fingers are not in position to bind dNTP (see above). Taken together, these considerations suggest that during the replication cycle of family B pols, there is concerted movement of the exonuclease, NH_2-terminal domains, and fingers relative to the catalytic region of the pols.

This concerted movement may be the structural basis for the functional coupling of polymerase and exonuclease domains, which is unique to the pol α family. In this family it is possible to generate site-directed mutations in one domain that exert an indirect, negative effect on the other (Bebe-Kimmel & Hamm, 1978; Abdus-Sattar et al., 1984). This contrasts with pol I pols like KF, where these activities are completely confined to their respective domains (Ollis et al., 1985).

Molecular basis of thermostability

Thermococcus sp. 9°N-7 grows at temperatures of 65–90°C, and its pol has a temperature optimum of 70–80°C (Perler et al., 1996). It has a half-life of 6-7

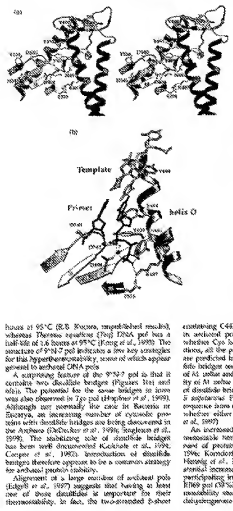


Figure 6. The active site of 9°N-7 pol and a modelled ternary complex for incorporation of the active site. Structures were evaluated individuals are discussed in the text. Hydrogen bonds are as broken lines, and the two disulfide bridges are shown as broken lines, and several ions are excluded as green spheres. The 9°N-7 pol and T7 pol ternary complex (Doublié et al., 1998) were superimposed on the palm (0.35 Å rmsd for 13 Cα atoms). The ternary conformation was adopted for D582 and D684 of 9°N-7 pol, and the β turn including D582 was then shortened, as a motion conjugate in was observed between the open/apo and binary complex structures of Bacillus stearothermophilus pol (Kiefer et al., 1998).

houses at 95°C (R.B. Roberts, unpublished results), whereas Thermus aquaticus (Taq) DNA pol has a half-life of 1.6 hours at 95°C (Perler et al., 1996). The structure of 9°N-7 pol indicates a few key strategies for this hyperthermostability, some of which appear general to archaeal DNA pols.

A surprising feature of the 9°N-7 pol is that it contains two disulfide bridges (Figures 3(a) and 4(b)). The potential for the same bridges to form was also observed in Tgo pol (Hopfner et al., 1999). Although not normally the case in Bacteria or Eucarya, an increasing number of cysteine proteins with disulfide bridges are being discovered in the Archaea (DeDecker et al., 1996; Stephens et al., 1998). The stabilizing role of disulfide bridges has been well documented (Kelehan et al., 1994; Cowan et al., 1992). Introduction of disulfide bridges therefore appears to be a common strategy for archaeal protein stability.

Alignment of a large number of archaeal pols (Edgell et al., 1997) suggests that having at least one of these disulfides is important for their thermostability. In fact, the two-stranded β-sheet containing C482 corresponds to sequence motif D in archaeal pols (Edgell et al., 1997). Based on whether Cys is present in the corresponding positions, all the pols observed by Edgell et al. (1997) are predicted to have at least one of the two disulfide bridges seen in 9°N-7 pol, with the exception of A1 pollur and B. stearothermophilus pol. The unavailability of A1 pollur pol may be partly caused by a lack of disulfide bridges. The B. stearothermophilus pol, like the 9 stearothermophilus P2 B3 pol, is highly divergent in sequence from other archaeal pols, and it is unclear whether either of these functions as pols (Edgell et al., 1997).

An increased number of salt-bridges relative to mesostable homologs is often cited as a determinant of protein thermostability (DeDecker et al., 1996; Karshikoff et al., 1995; Chan et al., 1998; Hennig et al., 1995). The 9°N-7 pol shows a substantial increase in the fraction of charged residues participating in salt-bridges (47 %) compared with B.7969 pol (30 %). These results are similar to a thermostability study of Pyrococcus furiosus glutamate dehydrogenase (Yip et al., 1995). The authors of

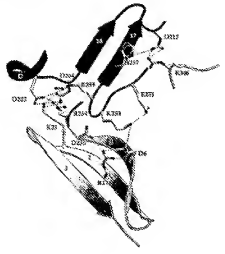


Figure 7. The extensive salt networks at the interface of the NH_2-terminal and 3'-5' exonuclease domains

that study found a marked preference for Arg residues in the ionic interactions of the thermostable enzyme, but no such preference is evident here. The same fraction (48 %) of Arg residues is used in ionic interactions in both 9°N-7 and B969 pols, whereas a much higher proportion of Glu residues participate in salt bridges in the 9°N-7 pol (83 %) compared with B969 pol (33 %).

The number and distribution of salt bridges within thermozyme does not substantially differ between 9°N-7 and B969 pols. As the interfaces between polydomains, however, the differences in ionic networks are striking. The proportion of ionic interactions at interfaces in the 9°N-7 pol (21 %) is over twice that in B969 pol (9%). The differences lie at the interface of the exonuclease domain with the NH_2-terminal domain (Figure 7), and at the interface of the exonuclease with the thumb, where a two and a three-membered ionic network occur in 9°N-7 pol compared with none in B969 pol (not shown).

Burial of the charged features of proteins has been cited as another factor that can confer thermostability (Hwang et al., 1995). The NH_2-terminal methione (M1) of 9°N-7 pol is stabilized by a hydrophobic cluster formed by L135, F327, I356, V358, and L141 while the corresponding residue of B969 pol is completely exposed to solvent. The β-factor for the C^{α} of M1 in 9°N-7 pol is 26 Å^2, whereas for M1 in B969 pol, it is 85 Å^2. While burial of the N terminus may be important for the thermostability of the 9°N-7 pol, the same does not hold for the C terminus. The last 25 residues are not visible in the electron density, similar to the case of B969 pol. The solvent accessibility of the C terminus of these pols may reflect the need for this region to interact with a processivity accessory protein, which is known to be the case in the T4 replication complex (Berdis et al., 1998).

Another common strategy for protein thermostability is to lower the solvent accessible surface area of the protein and to increase the proportion of buried structure (Kumchelwa et al., 1995, Chan et al., 1995). This translates into a more compact structural design. There are at least 11 examples of loop sequences in B969 pol that are much shorter or absent in 9°N-7 pol. Some of the more striking examples are shown in Figure 8. Alignment of 14 archaeal pols (Edgell et al., 1997) indicates that they share practically all of these sequence "deletions". The Tgo pol structure also revealed shortened loop segments relative to B969 pol (Hopfner et al., 1999). Nevertheless, the overall ratio of solvent-accessible surface area to volume for both 9°N-7 and B969 pols is the same (0.33). Thus, while lowering the surface area to volume ratio is a common strategy for thermostability, it is not the primary basis for the stability of 9°N-7 pol.

Materials and Methods

Purification, crystallization, and data collection

Thermococcus sp. 9°N-7 polymerase (wild-type and the D141A,D143A exonuclease-deficient mutant) was overexpressed and purified as described (Southworth *et al.*, 1999). Crystallization, cryoprotection, data collection, and reduction of native crystals are described (Zhao *et al.*, 1998). Derivatives were prepared by soaking native crystals in stabilization solutions (Zhao *et al.*, 1998) supplemented with 22.7 mM sodium ethylenediaminetetraacetate (Sigma-Aldrich) for 11 days (thymine-O2), 10 mM K_2PtCl_4 for one hour (PtCl-1), 1.5 mM dicyclohexylammonium-(ethylenediamine)-oxomolybdate (POM) for one day (POM-1), or 1.0 mM Baker's isomorphic for 30 hours (BaCl$_2$). These crystals were mapped through multiwavelength anomalous scattering (MAD) diffraction measurements. 10% (four molecules) and 20% acetone (one to two hours) additional derivatives were collected with the improved cryoprotection procedure reported (Zhao *et al.*, 1998) by soaking native crystals in 2.0 mM thymine-O2 for 6.5 days (thymine-O2). Six mM K_2PtCl_4 for seven days ($PtCl_4$), and 1.5 mM POM for 30 hours (POM).

Structure determination

The structure of the D141A,D143A mutant of 9°N-7 polymerase was determined by the method of multiple isomorphous replacement (MIR). A number of native and derivative crystals were used to solve the structure because of problems with anisomorphism (Table 1). Three native datasets were collected from single crystals: NAT-1 was mounted in the liquid nitrogen stream directly from cryoprotectant, whereas NAT-2 and -3 were flash-frozen in liquid nitrogen prior to mounting. The crystals belong to space group $P2_12_12_1$, with unit cell dimensions of approximately $a = 90.1$ Å, $b = 102.1$ Å, $c = 113.1$ Å, also NAT-3. One molecule is present per asymmetric unit, giving a solvent content of approximately 46%.

A difference Patterson map of thymine-O3 was calculated using the program MLPHARE in the CCP4 suite (CCP4, 1994). One heavy-atom site for this derivative was identified with the program RSPS (Knight, 1989). This site was used to calculate initial phases for NAT-1 at 5 Å resolution using the program MLPHARE (Otwinowski, 1991). Difference Fourier synthesis with the initial phases revealed three sites for the PtCl-1 derivative. Two native sets for this derivative were determined with the phases derived from both thymine-O 1 and PtCl-1. The correct handedness of the phasing information from these derivatives was determined using MLPHARE, and anomalous scattering data from the derivatives were included in the phase calculation. These sites for the BaCl$_2$ derivative and two other for POM were obtained from difference Fourier maps calculated to 5 Å resolution. All of these heavy-atom sites were included in subsequent phase calculations with NAT-1. The high-resolution phasing limit was extended to 3.5 Å. Because of the high solvent content in the crystals, use of the solvent-flattening program DM (Cowtan, 1994), in combination with histogram matching, improved the phases substantially. A polyalanine model was built into the improved electron density map of NAT-1 with the program O (Jones & Kjeldgaard, 1997) and refined in the program X-PLOR (Brünger, 1992). Phase combination using the program SIGMAA (Read, 1986) further improved the map during building and refinement.

Identification of side-chain densities was possible only after collecting a higher-resolution native dataset (NAT-2), along with difference data for three more derivatives obtained under improved cryoprotection conditions (thymine-O2, one site; $PtCl_4$, four sites; POM-2, five sites). These derivatives were used to calculate MIR phases at NAT-2 to 3.0 Å resolution. Partial model phases at NAT-2 were calculated using the refined polyalanine model derived from NAT-1. Because of significant differences in unit cell dimensions between NAT-1 and 2, it was not necessary to subject NAT-2 to rigid-body refinement against NAT-1 in X-PLOR. Combination of the polyalanine model phases and MIR phases with SIGMAA improved the electron density map. Model building, refinement, and phase combination were iterated until a complete polyalanine model could be built for the final stage of refinement. NAT-3 was used to extend the resolution limit to 2.8 Å, and water molecules were added.

Coordinate files and Illustrations

The *Thermococcus* sp. 9°N-7 polymerase atomic coordinates and structure factors have been deposited in the PDB Protein Data Bank under the accession code 3LDP. The RMS coordinate used for comparisons in this manuscript are those of the orthorhombic crystal form (coordinate code 1TAQ). Figures were prepared within the RIBBON program (Nielsen Graphics, Inc.) entirely (1B), 2, 3(a) and 3(b) or with images imported from MOLSCRIPT (1A-3 (Persson, 1991) or SETOR (Chi, 4-7) (Evans, 1993).

Acknowledgments

We thank New England Biolabs, Inc. for their collaboration on this project especially S. Ray Williams and Rebecca Kucera for protein purification and Francine Perler and William Jack for helpful discussions. We thank Abbie McCoy, Derek Olsen, Daniella Stack, Jean-Luc Jestin, and Hermann Bellinger for critical comments on the manuscript. We also thank Abbie McCoy, Jeff Stybar, and Sean Johnson for assistance in figure preparation. This work was supported by grants to L.S.B from the ACS (RPG-97-074, the March Caudron Biotechnology Career, and the Searle Scholars Program.

References

Abelson, Baker A. K., Lin, T. C., Jones, C. & Kumagawa, M. N. (1990). Functional correspondence and nonidentical kinetic parameters of prion proteins in bacteriophage T4 DNA polymerase. *Biochemistry*, **31**, 10817–10825.

Bessa, S. M., Dalrymple, C. K., Petersen, J. D. & Price, M. N. (1994). Perspectives on archaeal diversity, thermophily, and monophily from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA*, **91**, 9188–9193.

Brown, L. S. & Szalay, T. A. (1979). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 23–32.

Bernad, A. J., Soumillion, P. & Berkhout, B. J. (1990). The carboxyl terminus of the bacteriophage T4 DNA

polymerase is required for bacteriophage complementation. *Proc Natl Acad Sci USA* 85, 9268-9272.

Bernad, A., Blanco, L., Lazaro, J. M., Martin, G. & Salas, M. (1989) Conserved 3'-5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* 59, 219-228.

Bernstein, H. S. & Hu, J. (1994) Comparative sequence analysis and phylogenetic reconstruction of DNA polymerase motif A in vertebrates. *Nucleic Acids Res* 22, 2137-2143.

Bridgham, J. T. (1992). *X-Ray Structure of a Ternary Complex of DNA Polymerase I, DNA, and dTTP*. Yale University, New Haven, CT.

Brutlag, D. L., Atkinson, M. R., Setlow, P. & Kornberg, A. (1969) A thermolabile DNA polymerase from a mutant of *Escherichia coli*. *Biochem Biophys Res Commun* 37, 982-989.

Bryan, S. K., Bayle, J. H., Lampi, K. J., Bambara, R. A., Jessee, J. B. & Marshak, D. R. (1995) Characterization of DNA polymerase delta from bovine thymus. *Biochemistry* 34, 1041-1048.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol* 253, 333-346.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* 4, 89-96.

Bullock, P., Forster, K., Ostrow, J. D., Tsuchiya, Y. & Bambara, R. (1985) Fine structure mapping of mutations in the *Saccharomyces cerevisiae* DNA polymerase II gene. *Mol Cell Biol* 5, 1507-1516.

Burgers, P. M. & Bambara, R. A. (1979) Deoxyribonucleic acid (DNA) polymerase from *Saccharomyces cerevisiae*. A simple procedure for the purification of DNA polymerase I and DNA polymerase II to homogeneity. *J Biol Chem* 254, 7696-7701.

Burgers, P. M. & Kornberg, A. (1982) Mutagenic and lethal effects of the *dnaE*486 allele in *Escherichia coli* lacking exonuclease activity. *J Bacteriol* 150, 43-49.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment maturation in a dnaE486 mutant defective in DNA polymerase III holoenzyme. *J Biol Chem* 257, 10161-10166.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Accumulation of nascent DNA chains in polynucleotide ligase mutants. *J Biol Chem* 257, 10167-10173.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Joining of nascent DNA chains in polynucleotide ligase mutants. *J Biol Chem* 257, 10174-10178.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Processing of the 5'-terminus of nascent DNA chains. *J Biol Chem* 257, 10179-10183.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Processing of the 5'-terminus of nascent DNA chains. *J Biol Chem* 257, 10179-10183.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Joining of nascent DNA chains in polynucleotide ligase mutants. *J Biol Chem* 257, 10174-10178.

Burgers, P. M. & Kornberg, A. (1982) Mutagenic and lethal effects of the *dnaE*486 allele in *Escherichia coli* lacking exonuclease activity. *J Bacteriol* 150, 43-49.

Burgers, P. M. & Bambara, R. A. (1979) Deoxyribonucleic acid (DNA) polymerase from *Saccharomyces cerevisiae*. A simple procedure for the purification of DNA polymerase I and DNA polymerase II to homogeneity. *J Biol Chem* 254, 7696-7701.

Bullock, P., Forster, K., Ostrow, J. D., Tsuchiya, Y. & Bambara, R. (1985) Fine structure mapping of mutations in the *Saccharomyces cerevisiae* DNA polymerase II gene. *Mol Cell Biol* 5, 1507-1516.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol* 253, 333-346.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* 4, 89-96.

Bryan, S. K., Bayle, J. H., Lampi, K. J., Bambara, R. A., Jessee, J. B. & Marshak, D. R. (1995) Characterization of DNA polymerase delta from bovine thymus. *Biochemistry* 34, 1041-1048.

Bullock, P., Forster, K., Ostrow, J. D., Tsuchiya, Y. & Bambara, R. (1985) Fine structure mapping of mutations in the *Saccharomyces cerevisiae* DNA polymerase II gene. *Mol Cell Biol* 5, 1507-1516.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Accumulation of nascent DNA chains in polynucleotide ligase mutants. *J Biol Chem* 257, 10174-10178.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Processing of the 5'-terminus of nascent DNA chains. *J Biol Chem* 257, 10179-10183.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Joining of nascent DNA chains in polynucleotide ligase mutants. *J Biol Chem* 257, 10174-10178.

Burgers, P. M. & Kornberg, A. (1982) Mutagenic and lethal effects of the *dnaE*486 allele in *Escherichia coli* lacking exonuclease activity. *J Bacteriol* 150, 43-49.

Burgers, P. M. & Bambara, R. A. (1979) Deoxyribonucleic acid (DNA) polymerase from *Saccharomyces cerevisiae*. A simple procedure for the purification of DNA polymerase I and DNA polymerase II to homogeneity. *J Biol Chem* 254, 7696-7701.

Bullock, P., Forster, K., Ostrow, J. D., Tsuchiya, Y. & Bambara, R. (1985) Fine structure mapping of mutations in the *Saccharomyces cerevisiae* DNA polymerase II gene. *Mol Cell Biol* 5, 1507-1516.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol* 253, 333-346.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* 4, 89-96.

Bryan, S. K., Bayle, J. H., Lampi, K. J., Bambara, R. A., Jessee, J. B. & Marshak, D. R. (1995) Characterization of DNA polymerase delta from bovine thymus. *Biochemistry* 34, 1041-1048.

Bullock, P., Forster, K., Ostrow, J. D., Tsuchiya, Y. & Bambara, R. (1985) Fine structure mapping of mutations in the *Saccharomyces cerevisiae* DNA polymerase II gene. *Mol Cell Biol* 5, 1507-1516.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Accumulation of nascent DNA chains in polynucleotide ligase mutants. *J Biol Chem* 257, 10174-10178.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Processing of the 5'-terminus of nascent DNA chains. *J Biol Chem* 257, 10179-10183.

Burgers, P. M. & Kornberg, A. (1982) Okazaki fragment processing in a dnaE486 mutant. Joining of nascent DNA chains in polynucleotide ligase mutants. *J Biol Chem* 257, 10174-10178.

Burgers, P. M. & Kornberg, A. (1982) Mutagenic and lethal effects of the *dnaE*486 allele in *Escherichia coli* lacking exonuclease activity. *J Bacteriol* 150, 43-49.

Burgers, P. M. & Bambara, R. A. (1979) Deoxyribonucleic acid (DNA) polymerase from *Saccharomyces cerevisiae*. A simple procedure for the purification of DNA polymerase I and DNA polymerase II to homogeneity. *J Biol Chem* 254, 7696-7701.

Bullock, P., Forster, K., Ostrow, J. D., Tsuchiya, Y. & Bambara, R. (1985) Fine structure mapping of mutations in the *Saccharomyces cerevisiae* DNA polymerase II gene. *Mol Cell Biol* 5, 1507-1516.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol* 253, 333-346.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* 4, 89-96.

Kaster, L. K., Mayr, C., Brennan, J. C. & Beese, L. S. (1998) Structuring DNA replication in a radioactively active Bacillus DNA polymerase crystal. Nature, 394, 529–307

Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1992) Crystal structure of Thermus aquaticus DNA polymerase. Nature 376, 612–616

Korty, Bs., Brooks, R. D. & Jack, W. E. (1993) Characterization of a DNA polymerase from the hyperthermophile Archaea Thermococcus litoralis. J. Biol. Chem. 268, 9463–9471

Korolchuk, S., Hindge, E., Mishra, N., Rosenclay, A. & Jacobsen, H. (1996) The crystal structure of bacteriophosphodiesterase-3-phosphate dehydrogenase from the hyperthermophilic bacterium Thermotoga maritima at 2.2 A resolution. J. Mol. Biol. 260, 449–471

Kostelec, G., Akiyal, M., Bonner, W. M., Di Cera, E. & Thiaucourt, L. (1986) Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5 A resolution. Proc. Natl. Acad. Sci. USA, 92, 1024–1028

Li, Y., Korolev, S. & Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. 17, 7514–7525

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. Nature, 313, 762–766

Otwinowski, Z. (1991) In Isomorphous Replacement and Anomalous Scattering (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), pp. 80–86. Science and Engineering Research Council, Warrington, UK.

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. Science, 264, 1891–1903

Perler, F. B., Kumar, S. & Kong, H. (1996) Thermostable DNA polymerases. Advan. Protein Chem. 48, 377–435

Polesky, J. F. (1992) A paragon for practice: both detailed and substantial drawings for private structures. J. Appl. Crystallog. 24, 968–970

Rasolo, B. S. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallog. sect. A, 42, 140–146

Rossi, D. C. & Adams, M. W. W. (1992) Hyperthermophiles taking the heat and loving it. Structure 3, 331–334

Rubin-Bejerano, I. & Nancy, B. L. (1993). Genetic and biochemical studies of Saccharolobus 7d DNA polymerase 3'-5' exonuclease activity. J. Biol. Chem. 36, 27040–27048.

Rubinstein, I. E & Hurdnett, W. (1992). High-dose cytosine arabinoside in elevate hypothyroid industrie a chronic and pharmacologic antagon. Leuk. lymphomas 10, 43–48.

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nature Struct Biol. 4, 215–222

Sharrocks, L. C. & Bennett, P. P. (1989). Characterization of the L11, L1, L10, and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium Halobacterium salinobium. EMBO J. 8, 1325–1335.

Singleton, M. R., Scaglia, M. N. & Lindenbach, I. A. (1998) A new structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon Thermococcus litoralis. Structure, 6, 29–314

Southworth, M. W., Kong, H., Kucera, R. B., Ware, J., Jannasch, H. W. & Perler, F. B. (1996a) Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on Thermococcus sp. 9°N-7 and mutations affecting 3'-5' exonuclease activity. Proc Natl Acad Sci USA, 93, 5281–5285

Tahirov, S. & Richardson, C. C. (1995). A single nucleotide DNA polymerase of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc Nat Acad Sci USA, 92, 6209–6345

Terwilliger, T. C., Kim, S. H. & Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference patterson function. Acta Crystallog. sect. A, 43, 1–5

Thompson, J. F. & Bloomfield, C. D. (1979). A single nucleotide deletion in the polymerase gene of bacteriophage T7 is responsible for the small plaque mutant phenotype. Gene, 7, 261–275

Tsai, Y. C. & Richardson, C. C. (1988). Effects of Escherichia coli dnaQ gene mutations on DNA replication and viability of phage T7 in Escherichia coli. J. Bacteriology, 170, 5312–5317

Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H. & Steitz, T. A. (1996a). Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. Biochemistry, 35, 8110–8119

Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. Cell, 89, 1087–1099

Wheat, C. R., Sandle, O. & Wilhelm, M. L. (1996). Towards a rational system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci. USA, 87, 4576–4579

Wang, S. W., Wind, R. A. P., Yang, P. K., Azzi, N., Thomas, B. E., Aris, B., Scott, D. Thorbajarnarson, M. W. & Wang, T. S. F. (1998). Human DNA polymerase α gene expression is cell proliferation dependent and its promoter structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. EMBO J 7, 37–47

Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995) The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure 3, 1147–1158.

Zhou, M., Mou, C., Rodriguez, A. C., Korbo, J. R., Pisarcik, R. & Steele, L. S. (1998). Complications and preliminary determination analyses of a hyperthermostable DNA polymerase from a Thermococcus archaeon. Acta Crystallog. sect D, 54, 994–995.

Edited by D. Rees

(Received 23 June 1999; received in revised form 24 March 2000; accepted 24 March 2000)

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 278, No. 12, Issue of March 21, pp. 11024–11031, 2003
Printed in U.S.A.

Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase*

Received for publication, July 29, 2002, and in revised form, December 23, 2002
Published, JBC Papers in Press, December 26, 2002, DOI 10.1074/jbc.M208035200

Andrew F. Gardner‡, Catherine M. Joyce§, and William E. Jack§¶

From †New England Biolabs Inc., Beverly, Massachusetts 01915 and the §Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Comparative kinetic and structural analyses of a variety of polymerases have revealed both common and divergent elements of nucleotide discrimination. Although the parameters for dNTP incorporation by the hyperthermophilic archaeal Family B Vent DNA polymerase are similar to those previously derived for Family A and B DNA polymerases, parameters for analog incorporation reveal alternative strategies for discrimination by this enzyme. Discrimination against ribonucleotides was characterized by a decrease in the affinity of NTP binding and a lower rate of phosphoryl transfer, whereas discrimination against dideoxy was almost exclusively due to a slower rate of phosphodiester bond formation. Unlike Family A DNA polymerases, incorporation of 9-(2-hydroxyethoxymethyl)guanine triphosphate (where X is adenine, cytosine, guanine, or thymine) opposite T by Vent DNA polymerase was enhanced over dGTP by via a 50-fold increase in phosphoryl transfer rate. Furthermore, a mutant with increased propensity for nucleotide analog incorporation (Vent$^{(exo^-)}$ DNA polymerase) had unaltered dNTP incorporation while displaying enhanced nucleotide analog binding affinity and rates of phosphoryl transfer. Based on kinetic data and available structural information from other DNA polymerases, we propose active site models for dNTP, dideoxyTP, and acyclovirTP selection by hyperthermophilic archaeal DNA polymerases to rationalize structural and functional differences between polymerases.

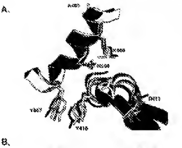
All free living organisms encode several DNA polymerases that are jointly responsible for the replication and maintenance of their genomes, thereby ensuring accurate transmission of genetic information (1–3). The majority of identified DNA polymerases can be classified into Families A, B, C, and Y according to amino acid sequence similarities to Escherichia coli polymerases I, II, III, and IV/V respectively (4, 5). Additional families have been identified, including the low-fidelity replicative DNA polymerases from hyperthermophilic Archaea (Family D) (6) and eukaryotic DNA polymerase η and terminal transferases (Family X) (4).

Structural and kinetic analyses of Family A (7–14) and Family B (15–20) DNA polymerases have increased the understanding of nucleotide selection and incorporation mechanisms. Although amino acid sequences diverge between these two families, the structures of Family A and B DNA polymerases share recognizable fingers, thumb, and palm subdomains. Also, these comparisons of structural elements important for function (7, 15). In the case of Family A DNA polymerases from bacteriophage T7, *Bacteriodes sp.* (Klenow fragment), large fragment of DNA polymerase I, and *Thermus aquaticus*, as well as the Family B DNA polymerases from bacteriophages RB69 interpretation of the structural information is complemented by nucleotide and pyrophosphate kinetic studies, allowing a detailed description of the polymerization pathway. Reaction parameters describing the discrimination against naturally occurring nucleotide analogs encountered in cells, such as NTPs, or antiviral nucleotide analogs, such as ddNTPs and dideoxy-

* This work was supported by National Institutes of Health Grant GM-29832. This paper is dedicated to the memory of Dr. Ira Schildkraut, who was a friend and colleague to us all.

¶ To whom correspondence should be addressed: New England Biolabs Inc., 32 Tozer Rd., Beverly, MA 01915. Tel.: 978-927-5054; Fax: 978-921-1350; E-mail: jack@neb.com.

A.



B.

	Region II	Region III
RB69	411 DLTSLYPSII 420	557 INRKLLINSLY 567
Vent	407 DFRSLYPSII 416	487 RAIKLLANSYY 497
9°N	404 DFRSLYPSII 413	485 RAIKILANSFY 495
KOD	404 DFRSLYPSII 413	484 RAIKILANSFY 495
TGO	404 DFRSLYPSII 413	484 RAIKILANSFY 495

FIG. 1. Alignment of Family B DNA polymerase active sites. A, [illegible]

[illegible]

EXPERIMENTAL PROCEDURES

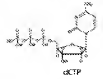
[illegible]

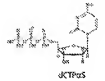
USSN 10/089,027

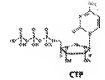
Atty. Docket No. 2007651-0001

4603542v1

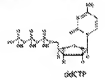
A. Nucleotides

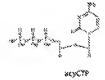




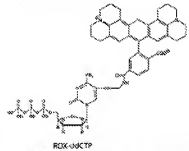


B. Nucleotide terminators





C. Dye-nucleotide terminators



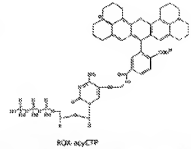


FIG. 2. Nucleotide and nucleotide analogs used for study of Vent DNA polymerase pre-steady-state kinetic mechanism. The maximum rate of nucleotide addition and the nucleotide constant for binding were determined with the following nucleotides and nucleotide analogs: dCTP, dCTPαS, ddCTP, acyCTP, (A) nucleotide terminators ddCTP and acyCTP (B), and dye-terminator nucleotides ROX-ddCTP and ROX-acyCTP (C).

constants or analog constants above, and found to the hyperbolic equation $k_{obs} = k_{pol}[nucleotide]/(K_d + [nucleotide])$, yielding k_{pol}, the maximum rate of nucleotide addition, and K_d, the dissociation constant for nucleotide binding (Eq. 1). The minimum energy difference between dCTP and nucleotide analog incorporation was calculated by Equation 1 (Eq. 1).

$$S(t) = -2[1 - \exp(-k_{obs}t)]/[enzyme]_o + k_{off}/k_{obs} + t \quad \text{(Eq. 3)}$$

Single-turnover kinetics were measured using manual mixing techniques. We established that in 10 μM Vent DNA polymerase was sufficient under the reaction conditions described by demonstrating that the rates of dCTP incorporation were the same using Vent DNA polymerase concentrations of 5 μM, 10 μM, and 20 μM and chose one of these.

Measurement of Pyrophosphorolysis Catalyzed by Vent DNA Polymerase—To measure the rate of DNA strand-directed pyrophosphorolysis by Vent or Vent^{exo-} DNA polymerases 50 nM enzyme was equilibrated with the DNA substrate (200 nM) in 1% Thermopol buffer and then mixed with PP_i in 1% Thermopol buffer at 60 °C using rapid quench techniques as described above. The extent of primer-strand cleavage at each time point was calculated by multiplying the cpm fraction of each DNA species by the number of phosphodiester bonds hydrolyzed to generate that species. k_{obs}, and k_{max} were derived using fitting protocols and equations as those described above for nucleotide addition.

RESULTS

Analysis of dNTP Incorporation by Vent DNA Polymerase—Previous studies with Family A DNA polymerases have shown that the steady-state pre-binding step for addition of a single correctly paired dNTP across phosphothioate bond terminus, i.e. I.3, 4A, 5G. Consequently, the first round of polymerization occurs more rapidly than subsequent rounds, resulting in a rapid "kick" burst of product. Incorporation of dCTP by Vent DNA polymerase displayed a burst pattern similar to those seen with RB69 exo Asp/Ala-618 DNA polymerase, with a rapid burst ($k_{pol,obs} = 62 \ s^{-1}$) followed by slow steady-state turnover ($k_{ss} = 0.05 \ s^{-1}$) (Fig. 2A and Table I). As indicated above, the burst is diagnostic for a rate-limiting step following bond formation, moreover, its amplitude is equal to the concentration of active enzyme, indicating that >90% of the Vent DNA polymerase preparation was active. Under similar conditions, Vent DNA polymerase failed to show a significant burst with ddCTP (Fig. 2B) or CTP (data not shown) incorporation. These data suggest that the rate-limiting step during nucleotide analog incorporation has changed compared with dNTP. Upon substitution of ddCTP with ddCTP-αS, both Vent and Vent^{exo-} DNA polymerases showed a 10- and 6-fold thio substrate effect ($k_{pol,obs}(dCTP)/k_{pol,obs}(ddCTP\text{-}\alpha S)$, respectively (Table I), consistent with an altered rate-limiting step.

Determinations of K_d and k_{pol} for dCTP addition by Vent DNA polymerase gave kinetic information similar to those determined for other DNA polymerases (Figs. 4A and Tables II and III). The relatively high K_d for nucleotide ($K_d \approx 70$ μM) is insensitive to the K_d for nucleotides determined in multiple turnover steady-state measurements ($K_m = 42$ μM) (3). Kinetic constants alone limit discrimination by nucleotide identity, as neither Vent DNA polymerase binding ($K_d = 26$ μM) and rate ($k_{pol} = 64 \ s^{-1}$) constants were observed for dATP incorporation. Substitution of dCTP with dCTP-αS had little effect on binding (K_d) or phosphodiester bond formation (k_{pol}); thus, the polymerase displays a maximum 10-fold discrimination against $k_{pol,dCTP}/k_{pol,dCTP\text{-}\alpha S} = 2.30$ (Table II).

Analysis of Vent DNA Polymerase-catalyzed Pyrophosphorolysis—To examine Vent DNA polymerase pyrophosphorolytic activity, we established dependence of a Fields-Isherwood enzyme kinetic display on the presence of increasing concentrations of PP_i. The dependence of the rate of Vent DNA polymerase pyrophosphorolysis on PP_i concentration yielded an equilibrium dissociation constant for PP_i binding of $K_d = 340$ μM and a maximum velocity of $k_{max} = 2.1 \ s^{-1}$ (Table IV).

Analysis of Ribonucleotide and Nucleotide Analog Incorporation by Vent DNA Polymerase—Kinetic parameters of ribonucleotide incorporation were determined to explore the effect of this precursor in 3′-OH ribonucleotide on polymerization. Vent DNA polymerase discriminated strongly against CTP incorporation with a 16-fold reduced binding affinity ($K_d \approx 1100$ μM) and a 410-fold slower rate of nucleotide addition ($k_{pol} = 0.155 \ s^{-1}$) (Table II). Comparisons of CTP and dCTP parameters interpreted as the ratio of catalytic efficiencies (k_{pol}/K_d)$_{dCTP}$/(k_{pol}/K_d)$_{CTP}$ revealed that Vent DNA polymerase preferred dCTP over CTP by 6600-fold.

In contrast to CTP, discrimination by Vent DNA polymerase against ddCTP and acyCTP was almost exclusively due to a slower rate of nucleotide addition, with K_d values for dCTP, ddCTP, and acyCTP being roughly equal (Fig. 4B and Table II). Indeed, the approximate 20-fold preference for acyCTP over ddCTP incorporation was almost entirely be attributed to steps associated by k_{pol}.

Similar experiments with Klenow fragment DNA polymerase also showed a 50,000-fold higher discrimination against acyCTP efficiency steps measured by both K_d and k_{pol}. The Klenow fragment DNA polymerase equilibrium binding constant for acyCTP was increased by 20-fold compared with dCTP and ddCTP, whereas k_{pol} for acyCTP incorporation was no-

FIG. 3. Pre-steady-state burst kinetics of dCTP and ddCTP incorporation by Vent DNA polymerase. Concentrations of enzyme-bound substrate (40 nM) in product (P) were determined by 5′-end-labeled substrate (40 nM) in product (P) were determined by 5′-end-labeled DNA polymerase with DNA (40 nM) and dCTP or ddCTP as indicated. Reactions were initiated by rapid mixing of enzyme-DNA complex with nucleotide and quenched at various times by addition of EDTA. Product concentrations were plotted as a function of time, and fit to the burst equation (equation 1, $A = -\exp^{-k_{obs}t} + k_{ss}t$). In this Vent DNA polymerase dCTP burst amplitude (A) was 35 nM, the burst rate (k_{obs}) was equal to 62 s^{-1}, and the steady-state rate (k_{ss}) was equal to 0.05 s^{-1}. In B the first-order rate constant for ddCTP incorporation was 0.6 s^{-1}.

Table I
Pre-steady-state burst kinetics

The kinetic parameters for Vent and Vent^{exo-} DNA polymerases are those of a two-nucleotide incorporation and are reported as the mean ± S.E. SD, not determined.

Enzyme	k_{max}, s^{-1}	k_{max}/K_d, μM^{-1} s^{-1}	k_{exo}, s^{-1}	k_{max}/k_{exo}
Vent	66 ± 49	0.20 ± 0.09	0.47 ± 0.08	0.003 ± 0.002
Vent^{exo-}	80 ± 5	0.19 ± 0.01	0.38 ± 0.02	0.005 ± 0.0003
KOD[a]	200 ± 99	2.2 ± 0.3	ND	ND
9°Nm[b]	30 ± 7	2.0 ± 0.2	ND	ND

[a] Ref. 31.
[b] Ref. 15.

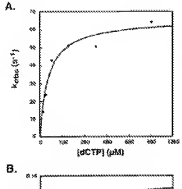


FIG. 4. Vent DNA polymerase pre-steady-state kinetics of nucleotide and nucleotide analog incorporation. The dependence of the observed rate (k_{obs}) on nucleotide or nucleotide analog concentration was fit to a hyperbola according to the Michaelis–Menten equation: $k_{obs} = (k_{max} \cdot [nucleotide])/(K_d + [nucleotide])$, where k_{max} is the observed rate of the nucleotide addition reaction, and K_d is the equilibrium dissociation constant for the nucleotide interaction. (*A*) Vent DNA polymerase (150 nM) was rapidly mixed with dCTP concentrations ranging from 0 to 500 μM and $k_{max} = 66 s^{-1}$, $K_d = 330 \mu M$ for the reaction. (*B*) Vent DNA polymerase (150 nM) was rapidly mixed with ddCTP concentrations ranging from 0 to 500 μM and $k_{max} = 37 s^{-1}$, $K_d = 190 \mu M$ for the reaction.

(except for ~2-fold) compared with dCTP (Table III).

ROX-ddCTP and ROX-acyclCTP Incorporation—Previous studies found ROX derivatized ddCTP and acyclCTP to be more efficient terminators than their unmodified forms when using Vent DNA polymerase (29). Pre-steady-state kinetics revealed higher binding affinities, but slower incorporation kinetics for the ROX derivatives (Table II), resulting in only marginal alterations in incorporation selectivity.

Analysis of Mechanistic Nucleotide Analog Incorporation by $9°Nm^{exo-}$ DNA Polymerase—We previously reported enhanced incorporation of nucleotide analogs by Vent^{exo-} DNA polymerase (15, 16). In a burst kinetics experiment, like Fig. 4*A*, mutant enzyme gave an initial burst of dCTP incorporation at a rate similar to that seen with the wild-type enzyme ($k_{max} = 45 s^{-1}$, >70% active) (Table I). Moreover, the single turnover kinetic parameters for dCTP addition ($K_d = 77 \mu M$ and $k_{pol} = 58 s^{-1}$) were similar to values derived for the wild-type enzyme (Table II). However, following the initial turnover, the steady-state rate of the A488L mutant polymerase was 5-fold slower than that of the wild-type enzyme ($k_{ss} = 8.39 s^{-1}$) (Table I), accounting for the lower specific activity of Vent^{exo-} DNA polymerase (32). As with wild-type Vent DNA polymerase, replacement of dCTP with dCTPαS had little effect on K_d or k_{pol} (Table II). Vent^{exo-} DNA polymerase was less active in phosphodiesterase playground with the wild-type enzyme, the result of a 3-fold reduction in PD, binding affinity and a 2-fold decrease in k_{pol} (Table IV). Incorporation of ITP, dGTP, aapCTP, and ROX-ddCTP by Vent^{exo-} DNA polymerase was more efficient, by 5–10-fold compared with incorporation by the wild-type enzyme, in each case, this is attributable to both increased binding affinity (lower K_d) and faster reaction rates (k_{pol}) (Table II). Incorporation of ROX-acyclCTP was largely unaffected by the A488L mutation (Table II).

DISCUSSION

The fundamental information derived for Vent DNA polymerase incorporation of dCTP confirms and expands earlier steady-state data (24) and places Vent DNA polymerase in full consort of other Family A and B DNA polymerases. As with these other polymerases, the steady-state rate for single nucleotide addition is limited by a slow step after phosphodiester bond formation. Previous steady-state measurements made use of a stock all four dCTPs were present gave a k_{cat} value of $12 s^{-1}$. This value is higher than the steady-state rate derived here by burst experiments ($0.9 s^{-1}$), most likely reflecting the higher concentrations (2.5 mM) and, more importantly, the processive requirements allowed by the earlier studies. In contrast, the experimental design reported here forces the DNA polymerases to act as a distributive enzyme, by dissociating from the DNA before binding another primer-template and incorporating an other nucleotide.

The single turnover parameters for Vent DNA polymerase with the natural dCTP substrate are similar to those of other Family A and B polymerases, both mesophilic and thermophilic. As shown in Table III, K_d and k_{pol} values define by

TABLE II

Pre-steady-state burst constants for nucleotide and nucleotide analog incorporation by Vent and Vent$_{(exo^-)}$ DNA polymerase

In all reactions, the kinetic parameters for Vent and Vent$_{(exo^-)}$ DNA polymerases are from at least two independent determinations (except where indicated). The constants are also reported as the average ± SD.

Nucleotide	Vent DNA polymerase				Vent$_{(exo^-)}$ DNA polymerase			
	K_d	k_{pol}	k_{pol}/K_d	Incorporation[a]	K_d	k_{pol}	k_{pol}/K_d	Efficiency[b]
	μM	s^{-1}	$s^{-1}\ \mu M^{-1}$		μM	s^{-1}	$s^{-1}\ \mu M^{-1}$	
dCTP	79 ± 7	68 ± 1	8.6×10^{-1}		72 ± 9	57 ± 5	7.9×10^{-1}	
ddCTP/ddA	1700 ± 90	0.11 ± 0.03	6.5×10^{-5}	1.4	63 ± 25	0.06 ± 0.01	9.5×10^{-4}	1.0
dTTP	1700 ± 200	0.160 ± 0.007	1.0×10^{-4}	0.042	260 ± 30	0.10 ± 0.01	3.8×10^{-4}	0.42
ddaTTP	80 ± 2	0.18 ± 0.01	2.3×10^{-3}	370	15 ± 0	0.05 ± 0.01	3.6×10^{-3}	42
ssaTTP	50 ± 20	7.6 ± 1.1	2.7×10^{-1}	10	20.5 ± 0.9	7.1 ± 0	3.4×10^{-1}	1.4
3OH-ddaTTP	35 ± 1	0.0059 ± 0.0010	1.7×10^{-4}	923	0.6[c]	0.02[c]	3.0×10^{-2}	70
2OH-ddaTTP	0.6 ± 0.2	2.08 ± 0.01	3.5×10^{0}	4	0[c]	1[c]	1.6×10^{0}	0.4

[a] Selectivity between incorporation of dCTP and other NTPs as the ratio of the efficiency of dCTP incorporation (k_{pol}/K_d) to the efficiency of the dddTTP, or ddGTP analog taken.

[b] The kinetic parameters for Vent$_{(exo^-)}$ DNA polymerase are from single determinations.

TABLE III

Pre-steady-state kinetic constants for nucleotide analog incorporation by RNA polymerases

The kinetic parameters for Vent and Vent$_{(exo^-)}$ DNA polymerases are from at least two independent determinations and are reported as the average ± S.D. All, not determined.

	dCTP		ddCTP		ssaTTP			
	K_d	k_{pol}	K_d	k_{pol}	K_d	k_{pol}	Efficiency	Selectivity[a]
	μM	s^{-1}	μM	s^{-1}	μM	s^{-1}	$s^{-1}\ \mu M^{-1}$	
Vent	79 ± 7	68 ± 1	63 ± 7	0.06 ± 0.01	260 ± 30	0.1 ± 0.01	3.8×10^{-4}	1
Vent$_{(exo^-)}$	72 ± 9	58 ± 5	63 ± 25	0.06 ± 0.01	20	0.2 ± 0.04	10×10^{-3}	26
9°N	80 ± 10[b]	105.6 ± 4.6[b]	80.5 ± 2.5[b]	0.08[b]	4500 ± 900[b]	0.13 ± 0.02[b]	29.0[b]	760
Klenow	9.5 ± 2.2	20 ± 1[c]	10.5 ± 1.7[c]	0.009[c]	8.4 ± 1.3	0.009[c]	1.0[c]	Insufficient data	0.026
Klenow$_{(exo^-)}$	10 ± 3[c]	7.1 ± 0.7[c]	9[c]	0.01[c]	10 ± 2.3	0.03 ± 0.01[c]	0.003[c]	8[c]	0.21

[a] Selectivity between incorporation of dCTP and other NTPs as the ratio of the efficiency of dCTP incorporation (k_{pol}/K_d) to the efficiency of the dddTTP, or ddGTP analog taken.

[b] Ref. 29.

[c] Ref. 35.

[d] Ref. 37.

[e] Ref. 19.

TABLE IV

Pyrophosphorolysis

The kinetic parameters for Vent and Vent$_{(exo^-)}$ DNA polymerases are from at least two independent determinations and are reported as the means ± S.D.

Enzyme	$K_{app(PP_i)}$	k_{pol}	$k_{pol}/K_{app(PP_i)}$	$k_{pol}/K_{app(PP_i)}$ with inhibitor[a]	$k_{pol}/K_{app(PP_i)}$ with inhibitor[b]
	μM	s^{-1}	$s^{-1}\ \mu M^{-1}$		
Vent	160 ± 160	1.10 ± 0.07	2.2×10^{-2}		300
Vent$_{(exo^-)}$	1000 ± 900	0.61 ± 0.31	0.7×10^{-2}		1200
RB69	10,000	0.01	1.0×10^{-6}		100,000
Klenow	1000	0.01	2.0×10^{-5}		6000

[a] A comparison of DNA polymerase activity with pyrophosphorothioate analogue is given by the ratio of DNA polymerase efficiency ($k_{cat}/K_{app(PP_i)}$) divided by the efficiency of pyrophosphorothioate ($k_{cat}/K_{app(PS_i)}$).

[b] Ref. 38.

[c] Ref. 6.

(3'-OH) shield for all polymerases tested, with no clear distinction between Family A and B DNA polymerases. Furthermore, the Family A Klenow fragment and Family B Vent and RB69 DNA polymerases carry out the reverse reaction of DNA polymerases, pyrophosphorolysis, with similar rates (k_{pol}), and Klenow fragment and Vent DNA polymerases attain comparable PP$_i$ binding constants ($K_{d(PP_i)}$) (Table IV). Similarities in nucleotide incorporation kinetics and active site structures underscore the evolution of DNA polymerases to efficiently carry out DNA replication and repair. Significant kinetic differences between the polymerases become apparent only when measuring nucleotide analog incorporation and aberrational effects, as detailed below.

Polymerase Interface—Despite a similar level of selectivity against NTPs, the discrimination is supplied almost exclusively by steric-zone resistance by k_{pol} for Klenow fragment DNA polymerase, whereas Vent DNA polymerase shows not only k_{pol} effects but also a 16-fold weaker ground state binding of the nucleotide. RB69 DNA polymerase also shows effects in both K_d and k_{pol}, attaining an even higher discrimination by virtue of a 200-fold weaker ground state binding. Discrimination against 3'-OH sans, an large part, been attributed to a steric clash between the 3'-OH and a conserved side chain in the polymerase active site (36, 37). The kinetic parameters suggest that the steric clash is first encountered in the ground state nucleotide binding by Vent exo RB69 DNA polymerases, but does not affect Klenow fragment DNA polymerase until the transition state of the reaction. This could occur for example, if the α_2 helix for Klenow fragment DNA polymerase instantly encounters bending prior to a conformational shift that engages the 3'-OH sensing machinery.

Diphosphorothioates—When incorporating ddGTP, RB69 DNA polymerase discriminates at the level of both K_d and k_{pol}. Discrimination by Vent, Klenow fragment, and 9°N DNA polymerases with a 2OH nucleic acid N-terminus defense (2OH DNA polymerases I) almost exclusively as see

steps monitored by k_{pol} and not those measured by K_D with Vent DNA polymerase showing less discrimination than the other two polymerases. This parallel behavior appears to reflect a similar lack of 3'-OH interference in ground state substrate binding modes than a conserved set of nucleotide contacts.

On the surface, the non-linearity in k_{pol} relative the dNTP concentration by Vent and Klenow fragment DNA polymerases (20) suggests similar discriminatory mechanisms for these two enzymes, a conclusion reinforced by the absence of an exonuclease subset with dNTPase acting either enzyme. The apparent superposition of the limb of an exponential effect with sublinear/unlimited dNTPs with Klenow fragment and Vent DNA polymerases is that a monothiophosphate preceding pyrophosphate bond formation is rate-limiting. Similarly, the lack of a significant chemistry effect for Klenow fragment DNA polymerase incorporation of dNTPαS vs. 1120 means that steps preceding phosphodiester bond formation continue to be rate-limiting for that enzyme. In contrast, the substantial effect noted for Vent and Vent$^{(exo-)}$ DNA polymerase incorporation of dNTPαSs is an indication that the chemistry of phosphodiester bond formation significantly influences the rate-limiting step for these polymerases.

The k_{pol} rates with both polymerases were significantly slower for dNTPs than for dNTPs. 5000- and 400-fold for Klenow fragment and Vent DNA polymerases, respectively in the case of Vent DNA polymerase, this result reflect at least a slowing of the chemical rate, whereas for Klenow fragment DNA polymerase, at least the rate of pre-chemical steps must be slowed. Thus, the pre-chemical rate for Vent DNA polymerase dNTP incorporation is at least 10-fold faster than comparable steps for Klenow fragment DNA polymerase.

Conserved amino acids positioned within active Family A or B DNA polymerases active sites probe for accurately base-paired substrates and concomitantly align phosphates into a geometry required for nucleophilic reaction. As conserved by Franklin et al. (31) and Yang et al. (28) in the RB69 DNA polymerase (among several interactions and by analogy, in the Vent DNA polymerase active site) (Fig. 1), the dNTP discrimination residue maintains a favorable 3'-exo sugar conformation. This conformation is maintained by hydrogen bonds between the 3'-OH and a main chain amide (corresponding to Vent DNA polymerase position 1120) and a non-bridging β phosphate oxygen (Fig. 5, A and B). Nucleotide α, β, and γ phosphates are further stabilized by direct or water-mediated hydrogen bonds with amino side chains (Fig. 5, A and B). The absence of the 3'-OH in dNTPs disrupts hydrogen bonding such that β phosphate (and main chain) amides, potentially increasing the activation energy required to orient the α phosphate for phosphoryl transfer (Fig. 5C). Indeed, the measured energetic difference between dNTP and dNTP incorporations (2.8 kcal mol^{-1}) is equivalent to that expected for the loss of at least two hydrogen bonds in the dNTP transition state (32).

Although the active site inserting nucleoside differs, as the Family A (Klenow fragment) DNA polymerase the dNTP 3'-OH contributes 97 kJ mol^{-1} to transition state stabilization, accounting for inefficient dNTP incorporation (27). This energy loss is counteracted by the closely related 3'-DNA polymerase active side by a hydroxyl group on Tyr^{766} (Klenow fragment DNA polymerase has Phe in the analogous position) that interposes a hydrogen bond to stabilize the dNTP α phosphate in the transition state, re-establishing a hydrogen bonding network similar to interactions formed by dNTP 12b. As a result, 3'-DNA polymerase selectivity between dNTP and dNTP is greatly reduced, as is the selectivity of the analogous Phe to Tyr mutation in both Klenow fragment and Taq DNA polymerases (33).

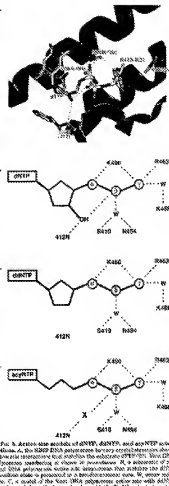


FIG. 5. Active site models of dNTP, dNTP, and ddNTP substrates. A, the RB69 DNA polymerase ternary complex (shown only with interactions that stabilize the substrate dNTP) (28). Vent DNA polymerase numbering is shown in parentheses. B, a schematic of the Vent DNA polymerase active site interactions that stabilize the dNTP transition state is presented in a pseudotetrahedral view. W, water molecule. C, a model of the Vent DNA polymerase active site with dNTP. Interactions that would be disrupted relative to dNTP are indicated. ddNTP (2', 3'-dideoxy-containing) nucleoside 12 a model for the binding of ddNTP suggests that, on the assumption of a similar ring, a nucleoside 12 which could be a similar nucleotide, in which the hydrogen-bonding between the Tyr^{766} main chain amide and non-bridging β phosphate

Nucleotide Selectivity. Divergence between polymerases is often noted upon acyNTP addition, again suggesting divergent modulation for nucleotide recognition and incorporation between the polymerases. Similar to dNTPs, acyNTPs lack the 3′-OH required to establish a hydrogen bonding network between the sugar, three metals of $Tp^{3/4}$, and a phosphate of the substrate (Fig. 5A). Klenow fragment DNA polymerase displays a strong discriminatory role on both K_d and k_{pol} terms, resulting in a selectivity value of 32,000 (in this case, the efficiency of acyNTP incorporation, or exactly as low as for an incorrect base pair $A_{template}$–$dGTP_{acyclic}$ = 5×10^{-4}, respectively; Table III) (16). A marked rate constant acyNTP incorporation bias also been noted for Taq DNA polymerase (24).

In contrast, acyNTPs are incorporated by hyperthermophilic archaeal DNA polymerases with only 10-fold lower efficiency than dNTPs by analogy with dNTP incorporation by Vent DNA polymerase, it seems reasonable that the space normally occupied by the sugar 3′- and 2′-carbons and associated solvation, would be intolerable to water molecules, metals, or protein side chains that might establish interactions to complements for those occupied by the missing 3′ OH groups. The difference in efficiencies from energy between dNTP and acyNTP incorporation ($\Delta\Delta G^{\ddagger}$ = $\Delta G^{\ddagger}_{acyclic}$ – $\Delta G^{\ddagger}_{deoxy}$ = 4.0 kcal/mol) is equivalent to a pair of two additional hydrogen bonds, which could be provided by hydrogen bonding between the main chain 112 amide, a putative water, and an acyNTP 3′-phosphate nonbridging oxygen, to induce interactions that exist in the dNTP active site (Fig. 5B). At the same time, we cannot rule out stabilizing interactions arising from residues near the active site potentially excluded by the ribose 2′- and 3′-carbons that are absent in acyNTP. Clearly, three-dimensional structural analysis will be necessary for a full understanding of the interactions important for Vent DNA polymerase incorporation of acyNTPs.

Base-substituted Nucleotides. Base-substituted nucleotides have been useful in a variety of analytical applications (40, 41). Not surprisingly, given the diversity of spin structures and charges, base-substituted nucleotides are accepted by DNA polymerases with varying efficiencies (28, 29, 35, 36). Previous studies identified nucleotide derivatives containing the deoxyribose ring or ribose as being more efficiently incorporated by Vent DNA polymerase than are canonical nucleotides lacking the dye (28, 35). On the current kinetic window, the magnitude of enhanced dye-substituted terminator additions was much lower than previously anticipated in nondenaturating gel fidelity analysis, even though some base pair intercalation energy was good agreement with the relative incorporation efficiency of dATP and acyNTP substitution (37). DNA substitutions of the nucleotide resulted in a 5–10-fold lower K_d, suggesting that contacts to or adjacent to the Vent DNA polymerase active site stabilize dye binding. However, in the same time, k_{pol} is reduced, suggesting that one consequence of the increased binding is to slow nucleotide addition. Thus, polymerase incorporation is a balance of both binding and catalysis: a construct bound with too high affinity requires higher activation energy for efficient turnover by the polymerase.

Vent^{exo-} DNA Polymerase Pre-steady-state Kinetics. Previous studies identified a Vent DNA polymerase variant (A488L) with enhanced nucleotide analog incorporation properties (38, 39). Current dNTP incorporation by Vent^{exo-} DNA polymerase is characterized by similar binding affinity (K_{d1}, nucleotide binding rate (k_{pol}), and rate-limiting step consistent with Vent DNA polymerase, presumably reflecting the conservation of residues actively involved in coordinating the incoming dNTP. In contrast, each of the nucleotide analogs tested with Vent^{exo-} DNA polymerase have higher binding affinity and faster rates of phosphoryl transfer than the unmodified polymerase. Energy differences between Vent and Vent^{exo-} DNA polymerase incorporation of dNTP or acyNTP are similar, $\Delta\Delta G^{\ddagger}_{dNTP}$ = 4.8 kcal mol^{-1} and $\Delta\Delta G^{\ddagger}_{acyclic}$ = 4.8 kcal mol^{-1}, suggesting that subtle hydrophobic or hydrogen bond-mediated effects could account for enhanced analog incorporation.

One hypothesis to account for these effects centralizes the A488L variant as being closer to the minimized conformational restriction facilitating incorporation of analogs. The residue analog given to Ala^{488} in the RB69 DNA polymerase crystal structures points away from the active site and base of the interface between the helix core of the polymerase and as a helix that must enable a 60° rotation to free the closed complex (Fig. 1A and 1B). DNA polymerases positioning a longer leucine residue at the position normally occupied by alanine in the α-helix may shift equilibrium from the open toward the closed conformation, thus reducing the activation energy for both binding and nucleotide transfer. This comes at a price: faster exonuclease discrimination that subsequent turnover by the A488L variant is inhibited, perhaps reflecting hindrance of the translation from closest to open states required for release and/or binding of the template and dNTP. This property does not, however, easily account for the fact that pre-steady-state kinetics for the nascent substrates are unaltered in this viscous chemistry. Resolution of this discrepancy may be on the genuine ability of this variant to overcome distortions in the nucleotide-binding site, distortions that are not present when the natural nucleotide is bound.

In summary, from these comparative studies, we observed that inclusion of dNTP incorporation pathways are conserved among Family B and B DNA polymerases despite diversity in primary amino acid sequences, macromolecular, function, and biological roles. However, differences in acyNTP and other nucleotide analog catalytic efficiencies by Klenow fragment, Vent, and other DNA polymerases illuminate fundamental differences underlying the kinetic pathway for DNA polymerization. As more DNA polymerases are studied functionally, it is apparent that motifs structural interactions in the active site influence how nucleotides are bound and positioned for catalysis.

Acknowledgments. We are grateful to Christie Hutchinson for answering this research as part of the Master of Natural Area Program at Fort Hunt University in A. P. G. We also thank Phil Buckley for providing the bacterial compounds and for helpful discussion, Alan Brownlie (Applied Biosystems) for useful discussions and technical advice, Hiroki Shibuta, Satoshi Iwahori, and Tom Evans for critical review of the manuscript, and Carla Bessett, Lucia Giacometti, and Lubo Rhoden for providing expert technical assistance. We are indebted to Dan Tawfik for fostering a supportive research environment at New England Biolabs Inc.

REFERENCES

1. Kornberg, A. (1980) *DNA Replication*, pp. 89–97, W. H. Freeman & Co., San Francisco.
2. Joyce, C. M., and Steitz, T. A. (1994) *Annu. Rev. Biochem.* 63, 777–822
3. Steitz, T. A. (1999) *J. Biol. Chem.* 274, 17395–17398
4. Doublié, S., Bi, K., and Ellenberger, T. (1998) *Structure Fold. Des.* 6, 799–808
5. Kim, J., Prasad, C., Rao, Lee, T., and Lawrence, C. (1995) *J. Mol. Biol.* 250, 549–587
6. Kunkel, T. A., Eastman, H., Pan, H., Sarah, R., and Mosbaugh, D. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 1494–1498
7. Patel, S. S., Wong, I., and Johnson, K. A. (1991) *Biochemistry* 30, 511–525
8. Patel, S. S., and Johnson, K. A. (1991) *Biochemistry* 30, 10602–10613
9. Doublié, S. S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* 391, 251–258
10. Doublié, S. S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* 391, 251–258
11. Steitz, T. A. (1993) *Curr. Opin. Struct. Biol.* 3, 31–38
12. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* 391, 251–258
13. Beese, L. S., Derbyshire, V., and Steitz, T. A. (1993) *Science* 260, 352–355
14. Johnson, K. A., Taylor, J. S., and Brown, L. R. (1982) *Proc. Natl. Acad. Sci. U. S. A.* 79, 3656–3660
15. Mildvan, A. S., and Fersht, A. R. (1977) *Biochem. Biophys. Res. Commun.* 80,

11862 *Vent DNA Polymerase Kinetics*

doi:10.1006/jmbi.2000.4455 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (2001) 305, 405–417

JMB

Crystal Structure of DNA Polymerase from Hyperthermophilic Archaeon *Pyrococcus kodakaraensis* KOD1

Hiroshi Hashimoto[1], Motomu Nishioka[2], Shinsuke Fujiwara[2] Masahiro Takagi[2], Tadayuki Imanaka[3], Tsuyoshi Inoue[1] and Yasushi Kai[1]*

[1]*Department of Materials Chemistry and*

[2]*Department of Biotechnology Graduate School of Engineering, Osaka University 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan*

[3]*Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University Yoshida-honmachi, Sakyo-ku Kyoto 606-8501, Japan*

*Corresponding author

The crystal structure of family B DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 (KOD DNA polymerase) was determined. KOD DNA polymerase exhibits the highest known exonuclease rate, processivity and fidelity. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanisms of these unique enzymatic features. Structural comparison of DNA polymerases from hyperthermophilic archaea highlighted the conformational differences in Thumb domains. The Thumb domain of KOD DNA polymerase shows an "opened" conformation. The fingers subdomain possessed many basic residues at the side of the polymerase active site. The residues are considered to be accessible to the incoming dNTP by electrostatic interactions. A 6-thiopurin motif (residues 242–249) existed from the Exonuclease (Exo) domain as seen in the editing complex of the RB69 DNA polymerase from bacteriophage RB69. Many aromatic residues are located at the forked-point like junction of the template-binding and editing clefts of KOD DNA polymerase, suggesting that the basic environment is suitable for partitioning of the primer and template DNA duplex and for stabilizing the partially melted DNA structure in the high-temperature environments. The stabilization of the melted DNA structure at the forked-point may be correlated with the high PCR performance of KOD DNA polymerase, which is due to low error rate, high elongation rate and processivity.

Keywords: archaea; crystal structure; family B DNA polymerase; "forked-point"; KOD DNA polymerase

Introduction

DNA polymerases are a group of enzymes that use single-stranded DNA as a template for the synthesis of the complementary DNA strand. These enzymes are multifunctional, with both synthetic (polymerase) and new or true degradative modes (3′-5′ and/or 5′-3′ exonucleases) and play an essential role in nucleic acid metabolism including the processes of DNA replication, repair and recombination. Many DNA polymerase genes have been cloned and sequenced. Amino acid sequences deduced from their nucleotide sequences can be classified into four major types. Escherichia coli DNA polymerase I (family A), *E. coli* DNA polymerase II (family B), *E. coli* DNA polymerase III (family C) and others (family X).[1] Recently, a new family of DNA polymerases has been identified, all members of this family contain five highly conserved motifs, I–V, and several of these polymerases participate in base repair.[2] This family is called the Din-rec/ExoIII family.[3] Family B DNA polymerases include eukaryotic DNA polymerases α, δ, and ε, which are thought to be components of the replisome and to carry out chromosomal DNA replication. Archaeal proteins involved in gene expression, such as those for DNA replication, transcription, and translation, have been found to be similar to those from eucarya. Therefore, the archaeal system of gene expression is a simplified model of the eukaryotic system. In contrast, the

* E-mail address of the corresponding author: kai@chem.eng.osaka-u.ac.jp

0022-2836/01/020405–13 $35.00/0

cellular appearance and organization of archaea are more like those of bacteria.

The first crystal structure of a family B DNA polymerase to be obtained was that of bacteriophage RB69 DNA polymerase (RB69 DNA polymerase).[7] The first crystal structure of archaeal DNA polymerase was DNA polymerase from *Thermatoga maritima* (Tma DNA polymerase).[8] The editing complex of RB69 DNA polymerase has been reported,[9] but further crystal structures of archaeal family B DNA polymerases have recently been reported: Tma DNA polymerase from *Thermococcus* sp. 9°N-7 DNA polymerase from *Thermococcus* sp. 9°N-7.[10]

The *Pyrococcus kodakaraensis* KOD1 is a hyperthermophilic archaeon, with an optimum growth temperature of 95°C.[11] Enzymes produced in KOD1 were reported to be extremely thermostable and to have eukaryotic characteristics.[12] The optimum temperature of KOD DNA polymerase is 75°C, similar to that of DNA polymerase obtained from *Pyrococcus furiosus* (Pfu DNA polymerase). KOD DNA polymerase, however, exhibits the largest extension rate (100–130 nucleotides/second) and processivity (>2 kb/min) five times and ten to 15 times higher than those of Pfu DNA polymerase, respectively.[13] Thermostable DNA polymerases are expected to be available enzymes for Polymerase Chain Reaction (PCR). KOD DNA polymerase is, therefore, suitable for DNA amplification by such means. Indeed, KOD DNA polymerase is widely used in rapid and accurate PCR systems (TOYOBO Ltd., Japan).

Although structures of three archaeal DNA polymerases have been determined as described above, no structural information relating to exonuclease activity, processivity or fidelity is provided. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanism of enzymatic features of KOD DNA polymerase, which are the highest extension rate, processivity and fidelity. Here, we report the crystal structure of DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. The three-dimensional structure of the KOD DNA polymerase may provide useful information to clarify the mechanisms for rapid and accurate synthesis. In addition, this information may contribute to the improvement of the PCR properties of enzymes already in use such as thermostability, error rate, elongation rate and processivity, or for designing new enzymes for PCR as well as DNA replication by family B DNA polymerases.

Results and Discussion

Overall structure

KOD DNA polymerase has a dish-like shape with dimensions of 65 Å × 65 Å × 102 Å and is made up of distinct domains and subdomains N-terminal (Ntase 1–108, 307–366, violet), Exonuclease (Exo 133–306, blue), Polymerase (Pol) domain including the Palm and Fingers subdomains (309–449, 560–697, brown, and 450–559, green, respectively) and the Thumb domain including thumb-1 and thumb-2 subdomains (698–774, red) (Figure 1(a)). The polymerase active site, containing three conserved carboxylates, Asp-404, Asp-540 and Asp-542) is located in an anti-parallel β-sheet in the Palm subdomain. The exonuclease active site contains two conserved carboxylates (Asp-141 and Glu-143) and is located in an anti-parallel β-sheet in the Exo domain. The Polymerase and exonuclease active sites are the molecular surface are indicated by P and E, respectively (see Figure 1). Structural components of archaeal DNA polymerases (KOD, Tma and 9°N-7 DNA polymerases) are shown in Figure 1(b). The structural architectures of the proteins are identical, but the orientations of the domains and subdomains is different. In the case of the KOD DNA polymerase (red), the Thumb domain is shifted to make an "open" conformation and the position of the Palm domain neighboring the root of the Thumb domain is slightly shifted as a result of the large movements of the Thumb domain in comparison to other archaeal DNA polymerases. Table 1 shows the averaged temperature factors of the domains and subdomains in the crystal structure of KOD DNA polymerase. The value of the Thumb domain was markedly higher than the others. The structures of many residues in the Thumb-2 subdomain are not defined, because the connection of the subdomain is highly disordered. Therefore, it is thought that the structure of KOD DNA polymerase described here provides information for the DNA-free most relaxed conformation. The structure of the editing complex of RB69 DNA polymerase revealed that newly synthesized duplex DNA is grasped by the Pol and Thumb domains. Although the orientation of the Thumb domain is potentially highly flexible, the orientation may be fixed when it binds to the primer-template duplex.

Polymerase domain

The Pol domain is made up of the Fingers and Palm subdomains and has an "L-like" shape (Figure 2(a)). The polymerization mechanism has been studied mainly on family A DNA polymerases (Pol-I). A structural basis for a motif

Table 1 Averaged temperature factors

Domain	Temperature factor (Å²)
Ntase	58.1
Exo	70.7
Pol	
Fingers	49.9
Palm	52.8
Thumb	85.7
Overall	59.9

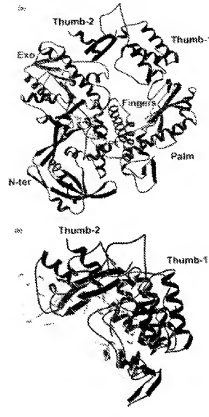


Figure 1. (a) Overall structure of KOD DNA polymerase. The structure is composed of domains and subdomains which are N-terminal (blue), exo (cyan), Palm (green), Fingers (green), subdomain 2 (green), Thumb-1 (red), Thumb-2 (red), and N-ter (yellow). Conserved carboxylic residues in Palm and exo domains are shown by ball-and-stick models. (b) Conformational comparison of Thumb domains among three archaeal DNA polymerases. Red, KOD DNA polymerase; blue, Tgo DNA polymerase; and green, *9°N-7* DNA polymerase. The comparison shows that the Thumb domain of KOD DNA polymerase displays the most "opened" conformation.

assisted mechanism of phosphoryl transfer was presented by the bacteriophage T7 DNA replication complex.[1] The complex structure shows that two metal ions are located in strictly conserved carboxylates (Asp475 and Asp654, which correspond to Asp501 and Asp648 in KOD DNA polymerase) extended from two anti-parallel β-sheet of the Palm domain. The phosphate group of incoming dATP is held by the metal ions and the four basic residues extending from the Fingers subdomain (Lys486, Arg518 and Lys522). The crystal structure of two tertiary complexes of the large fragment of

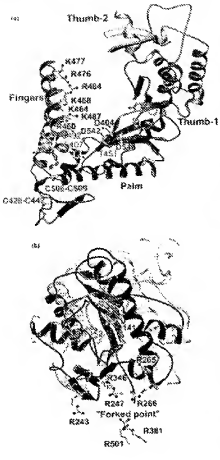


Figure 3. (a) Ribbon representation of the Pol domain. The domain is made up of Fingers and Palm subdomains. Conserved carboxylate residues (D404, D548 and E743) are concentrated by hairpin-rich models, hence members are represented by hairpin-rich models, which stand as a lane in the Fingers subdomain, facing the polymerase active site. Motifs are replaced by alleviate, because of the ambiguity of its electron density. Two disulfide bonds are displayed (C408–C442 and C505–C509). Aromatic residues adjacent to a glycine residue, represented by ball-and-stick models, are localized in the lobes of the subdomains. The Thumb domain is represented by the semitransparent ribbon. The C^{α} atoms of the methionine residues, M37 (the Pol-I), M412 (in Pol-α) and T541 (in Pol-β), are represented by sticks spheres. (b) Ribbon rendered drawing of KOD DNA polymerase and other DNA polymerase homologues are evaluated. Conserved carboxylate (D141 and E143) and aspartic residues (R243, R245, R247, R266, R381, R501) in the forked point of RB69 DNA polymerase are represented by ball-and-stick models. Conserved aromatic residues (Y409 and F411) are shown as ball-and-stick models, respectively.

Thermus aquaticus DNA polymerase I (Klentaq) with a primer-template DNA and ddCTP have been reported.[4] The ternary complexes suggest that basic residues at the fingers subdomain bind the phosphate group of the incoming dNTP and the dNTP induces a conformational change to deliver the incoming nucleotide to the active site. In the case of family B DNA polymerases, the Herpes simplex virus is composed mainly of two long helices and does not have a palm that appeared in the structures of family A DNA polymerases. Therefore, it seems that in the case of archaeal DNA polymerases, the mechanism of the Pol domain to deliver dNTP to the active site differs from that of family A DNA polymerases. Kinetic studies of RB69 DNA polymerase mutants revealed that three residues (Arg482, Lys486, Lys580 and Asn584) of the Fingers subdomain affected dNTP incorporation.[5] The residues are conserved in family B DNA polymerases, and corresponded to Arg461, Lys465, Lys487 and Asn491 in 9ºN DNA polymerase, respectively. Furthermore, Lys580, Arg476, Lys477 and Arg484 are located at the tip of the Fingers subdomain on the side of the polymerase active site in RB69 DNA polymerase (Figure 2b)[5]. It is expected that the "gauze" of basic residues captures the incoming dNTPs, then the dNTP is delivered toward the polymerase active-site center by accompanying the movement of the polymerase domain. Two disulfide bonds exist in the connection site between the Palm and Fingers subdomains (Figure 2a): Cys324-Cys452 and Cys506-Cys509). The two disulfide bonds are found also in the crystal structures of Tgo Pol and 9ºN-7 DNA polymerases. Sequence alignments for archaeal DNA polymerases is shown in Figure 3, suggesting the potential for the formation of disulfide bonds in the same sites. It is thought that the disulfide bonds are required to maintain the structure of the Fingers and Palm subdomains at extremely high temperatures. Sequence comparison suggests that the number of disulfide bonds are correlated with optimum growth temperatures of organisms. DNA polymerases from *Thermococcus litoralis*, *Methanococcus jannaschii* and *Archaeoglobus fulgidus*, with optimum growth temperatures at 85, 85 and 85°C, respectively, are expected to have one disulfide bond, because Cys486 is replaced by serine in *T. litoralis* and *M. jannaschii*, and Cys442 is replaced by arginine in *A. fulgidus*. DNA polymerase from *Methanobacterium thermoautotrophicum*, with an optimum growth temperature of 65°C, is expected to have no disulfide bond, because Cys438, Cys442 and Cys506 are replaced by glutamic acid, arginine and serine, respectively.

Archaeal DNA polymerases have characteristic sequences of aromatic residues adjacent to glycine residues (Figure 3). These are located at the hinges of the Palm subdomain at the connections to the Fingers and Thumb-2 subdomains (Figure 2b). These aromatic residues may provide a flexible structure, since aromatic bases of the adjoining glycine residues. This may contribute to the conformational changes of Pol domain in polymerization.

The 3'-5' exonuclease domain

DNA is synthesized by competition between the rate of polymerase and exonuclease activities at the newly synthesized 3' terminus from the primer. Misincorporation of a nucleotide destabilizes the structure of duplex DNA at the 3' terminus of the primer. This decreases the rate of nucleotide attack on the triphosphate group of the incoming dNTP by the primer 3'-OH, and allows excision of the incorrect nucleotide by the proofreading exonuclease. The complex requires the movement of the 3' terminus to the exonuclease active site accompanied by rewinding of the duplex DNA, because the exonuclease active site is set apart from the polymerase active site in RB69 DNA polymerase, the exonuclease active site is set apart from the polymerase active site by approximately 40 Å. The editing complex of RB69 DNA polymerase shows structural similarity to the editing mode of family B DNA polymerase.[6] The DNA polymerase binds the mismatched primer-template DNA, which is partially denatured, the 3' end of the primer strand is located at the exonuclease site. Residues 253-262 of RB69 DNA polymerase, that form an extended β-hairpin structure that juts directly out from the protein surface and projects into the DNA, destabilize the partially denatured or melted structure. Arg260 extending from the β-hairpin motif plays an important role. Arg260 and Phe123 appear to block the template strand by making interactions with the penultimate base at the 3' end of the primer-template. Arg260 and Phe123 in RB69 DNA polymerase correspond to Arg287 and Phe182, in RB69 DNA polymerase respectively. Figure 2(b) shows the structural comparison of Exo domains of RB69 and RB69 DNA polymerases. Molecular surface and ribbon-trace potentials are shown in Figure 4. The β-hairpin motif in RB69 DNA polymerase corresponds to residues 282-289 and Arg287, extending to the forked point, which is the junction of the template-binding and editing cleft (T-cleft and E-cleft, respectively) (Figure 4a, b) means that Arg287 can maintain template strand from primer strand and stabilize the melted structure of the strands in a manner similar to that of the RB69 DNA polymerase. As Phe182 is set apart from the active site, it is apparently unable to make an aromatic interaction with the base of the primer. Based on the above idea, the movement of the loop including Phe182 (Figure 2b) is expected to interact with the primer bound at the E-cleft. Furthermore, Arg247 extends from the β-hairpin structure to the T-cleft. Arg247 interacts with the template strand to lie it at the T-cleft. It is similar to Arg260 and Arg287. Five arginine residues gather at the forked-point in RB69 DNA polymerase (Arg286, Arg288, Arg346, Arg383 and Arg385) and provide a basic environment (Figures 2(b) and 4). It seems that they can interact with the phosphate

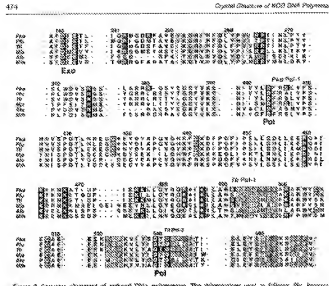


Figure 3. Sequence alignment of archaeal DNA polymerases. The abbreviations used in Fig. 2 are: *Pa*, *Pyrococcus kodakaraensis*; *Pw*, *Pyrococcus furiosus*; *Tk*, *Thermococcus litoralis*; *Mja*, *Methanococcus jannaschii*; *Aph*, *Archaeoglobus fulgidus*; and *Mth*, *Methanobacterium thermoautotrophicum*. Homologous residues are stacked in gray. Remarkable residues are highlighted in reverse type. Conserved carboxylate residues in the Exonuclease and Polymerase active sites are shown in red. Basic residues gathering at the forked-helix and Fingers subdomains are shown in blue. R103, R207, R208, R260, R266, R491 and R561 are located at the forked-point. R409, R464, R503, R460, R467, R469 and R480 are located in the Fingers' subdomain and face into the polymerase active site. Cysteine residues forming the possibly reversible disulfide bonds are shown in green. Hydrophobic residues in self-splicing inteins are shown in violet. Intein-host junctions both at the endonuclease residues. Antigenic sensitive segments to glycerol are shown in orange.

groups of the DNA strand and stabilize the melted structure of DNA strands at the forked-point. Several arginine residues at the forked-point are encountered in known family B DNA polymerases from hyperthermophilic archaea.

In DNA synthesis, the structure of DNA is variable at the stage of switching between the elongation- and melting modes. Hyperthermophiles must have mechanisms to protect their genomic DNA against thermal denaturations. The genomic DNA of hyperthermophilic archaea have nucleosome-like structure brought about by interaction with histone-like proteins[14] Hawarthalaea, as the replicative fork, the DNA strands are exposed. Therefore, DNA polymerases of hyperthermophilic archaea are required to stabilize the segment of melted DNA structure at the high temperature environment. The substitutions by DNA polymerase may correlate with the enzymatic characteristics of DNA polymerase such as fork-Rh period of activity, error rate, elongation rate and processivity. As discussed above, it is considered that the arginine residues around the "forked-point" have a remarkable effect on the stability of DNA structure. In the forked-point of Pfu DNA polymerase, Arg247, Arg345 and Arg581 are replaced by methionine, threonine and lysine, respectively. Therefore, the replacements may affect the difference of the enzymatic characteristics between RB69 and Pfu DNA polymerases. Additional experiments such as site-directed mutagenesis, a together with enzymatic studies of DNA polymerases are necessary to clarify the role of the residues at the forked-point.

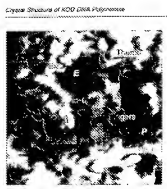


Figure 4. Molecular surface with electrostatic potential map around the bound palm. The red and blue surfaces are acidic and basic regions, respectively. Domains and subdomains are labeled with orange letters. Polymerase and thumb-base active sites are labeled with P and E, respectively. The β-hairpin is labeled with β.

Exonuclease site

The KOD DNA polymerase gene encodes a 3′→5′ exonuclease and nuclease processor protein. The precursor protein is processed precisely into these parts by protein splicing. The self-splicing reaction yields the mature KOD polymerase (775 residues) and two intervening protein elements (internal protein, Pfu-Pol-2ORF residues) and Pfu-Pol-2ORF residues as a result of the ligation of the external N and C-terminal domains (internal protein).[6][7] All known precursor proteins contain cysteine/serine amino acids at self-splicing sites: serine, threonine or cysteine (hydroxyl-amino) at the intein N terminus, and His-Asn pair at the intein C terminus followed by serine, threonine, cysteine (nucleophile) at the C-extein N-terminus.[8] The issues of the protein splicing reaction in KOD DNA polymerase are Ser407 and Ser410, which were located in at the N terminus of the C-extein in the crystal structure of KOD DNA polymerase, the nucleophilic residues are found in the Pol-domain (Figure 1(b)).

Subtopology rules in archaeal family B DNA polymerases in family are described into three types: Pfu Pol-1, Th Pol-3 and Th Pol-2 (The basic database, http://www.rub.com/~meh/aminohtml). The nucleophilic residues, serine or threonine, in the three sites are mapped in Figure 1(b). In the case of KOD DNA polymerase, Pol-Pol intervenes in the Pol Pol-1 site and Pfu-Pol2 intervenes in the Th Pol-1 site. The structure shows that they are localized around the polymerase active site in the Palm domain. Although they are exposed to solvent, they are surrounded by the fingers subdomain and the Thumb domain. The two loci(s) connect exist in the space because of steric hindrance. Therefore, it is necessary that the folding of intein and the subsequent self excision are carried out before the extein is folded.

Materials and Methods

Crystallization

KOD DNA polymerase was overexpressed in *E. coli* BL21(DE3) and purified by the previously reported method.[6] The crystals of KOD DNA polymerase were grown by the previously reported method.[17] KOD DNA polymerase was concentrated up to about an $A_{280\ nm}$ of 25. Crystals of KOD DNA polymerase suitable for diffraction experiments were obtained at 293 K with hanging drops of 2 μl of protein solution and 2 μl of reservoir solution containing 100 mM sodium citrate buffer (pH 5.6 and 25–30% (v/v) 2-methyl-2,4-pentanediol (MPD) equilibrated against the reservoir solution.

Data collection

X-ray diffraction measurements were performed at the beamline 18B of the Photon Factory at the High Energy Accelerator Research Organization, Tsukuba, Ibaraki City, Japan. Each crystal of KOD DNA polymerase was picked up directly with a nylon fiber loop from a drop of mother liquid, the crystal was then rapidly transferred to the N_2 gas stream. The beamline beam with wavelength of 1.00 Å was collimated to 0.2 mm at detector. Intensity data were collected on 205 mm × 420 mm imaging plates (Fuji Film Company Ltd) using the Weissenberg camera for macromolecules with a distance of 420 mm[18] and the oscillation method with 3° oscillation per frame. The crystals diffracted to least to 2.8 Å resolution at 100 K. X-ray diffraction data were processed and scaled with programs DENZO and SCALEPACK.[19] The diffraction data were scaled with even a (1/sin²θ) linear-rod parameters were determined as $a = 131.9$ Å, $b = 177.4$ Å, and $c = 74.0$ Å, with the space group of $P2_12_12_1$. The asymmetric parameters gave Matthews coefficient (V_m)[20] and a solvent content of 54.6% (v/v).[20] The final merging statistics of the data consisted of 99,587 measurements of 38,288 unique observed reflections with an overall R_{sym} of 6.4% and 94.5% in the outermost resolution shell (3.96–3.80 Å). The symmetry 98.1% in theoretically symmetric reflections and 2.8 Å resolution. The outermost resolution shell of data is 65.7% complete.

Structure determination

The crystal structure of KOD DNA polymerase was solved by molecular replacement with the AMoRe program.[21] The structure of the DNA polymerase (pdb code 1TGO) reduced to polyalanine was used as the search model. Data in the resolution range of 20.0–3.5 Å were used in the rotation and translation functions. Positional and temperature factor refinements of the AMoRe correlation coefficient (CC). Using a Fourier cut-off radius of 10 Å, a list of 25 solution families peaks was obtained, with the top peak having an AMoRe CC value of 19.8. The top solution by translation function is CC of 13.7 with an R factor of 54.1%. At this stage, the electron density of the Thumb domain is very ambiguous. Therefore, structure refinements of the initial stage was carried out with

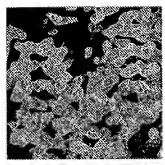


Figure 8. The final $2F_o - F_c$ map around the Phe410 and Tyr409 side-chains. The map is contoured at 1 σ.

a model covering the Thumb domain. The model was manually rebuilt using the program $O^{[24]}$ and subjected to further rounds of refinement using data in the resolution range 40.0–3.0 Å with the program CNS.[25] The final R-factor is 23.1% and R_{free} is 31.3% with rms deviations for bond lengths and bond angles being 0.007 Å and 1.3°, respectively. The 28 residues at the tip of one Thumb domain are not included in the final model due to poorly defined electron density. Figure 8 shows the final $2F_o - F_c$ map superimposed on the refined final coordinates of KOD DNA polymerase.

Protein Data Bank accession code

Refined coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the accession code 1JCL.

Figure preparation

Figures 1 and 2 were prepared using programs MOLSCRIPT[26] and RASTER3D.[27] Figure 3 was prepared by CHORD.[28] Figure 5 was prepared using the program GRASP.[29]

Acknowledgments

We thank Professor H. Sakabe, Dr N. Watanabe, Dr M. Suzuki and Dr Y. Igarashi, for support in data collection at KEK/PF, Japan. This study was supported by TARA Sakabe Project at University of Tsukuba. The author is grateful for a JSPS Fellowship for Japanese junior scientists.

References

1. Ito, J. & Braithwaite, D. K. (1991). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.* **19**, 4045–4057.
2. Ishino, Y., Komori, K., Cann, I. K. O., Koga, Y. & Kuramitsu, S. (1998). A specific mutation alters nucleotide discrimination and DNA synthesis activities in the B-family DNA polymerase from *Pyrococcus furiosus*. *J. Biol. Chem.* **273**, 8069–8075.
3. Braithwaite, D. K. & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucl. Acids Res.* **21**, 787–802.
4. Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. *Cell*, **89**, 1087–1099.
5. Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl Acad. Sci. USA*, **96**, 3600–3605.
6. Shamoo, Y. & Steitz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell*, **99**, 155–166.
7. Zhou, Y., Jeruzalmi, D., Steitz, J., Leighton, L., Lasken, R. & Kuriyan, J. (1998). Crystal structure of an archaebacterial DNA polymerase. *Structure*, **7**, 1289–1299.
8. Rodriguez, A. C., Park, H.-W., Mao, C. & Beese, L. S. (2000). Crystal structure of a pol α family DNA polymerase from the hyperthermophilic archaeon Thermococcus sp. 9°N-7. *J. Mol. Biol.* **299**, 447–462.
9. Hopfner, K. Eichinger, S. & Lipscomb, T. (1999). The mutant of archaeal geminin analysis, correlation and thermostably conferred. *Gene*, **476**, 113–119.
10. Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. & Imanaka, T. (1997). Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* **63**, 4504–4510.
11. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, **391**, 251–258.
12. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514–7525.
13. Yang, G., Franklin, M. J., Li, J., Konigsberg, W. H. (1998). Structure-based kinetic characterization of fd DNA polymerase mutants affecting the dNTP triphosphate recognition. *Biochemistry*, **34**, 8305–8316.
14. Hopfner, K. P., Eichinger, A. & Rieger, J. P. (1999). DNA stability and DNA binding proteins. *Mutat. Protein. Chem.* **48**, 437–467.
15. Nishioka, M., Fujiwara, S., Takagi, M. & Imanaka, T. (2000). Characterization of two amino terminal endonucleases encoded on the B700 polymerase gene of *Pyrococcus kodakaraensis* strain KOD1. *Mol. Gen. Genet.* **263**, 44–53.
16. Perler, F. B., Davis, E. O., Dean, G. E. & Adam, E. (1992). Comparison and analysis of intein sequences. *Nucl. Acids Res.* **20**, 3097–3099.
17. Kitabayashi, H., Nishimura, T., Nishioka, M., Fujiwara, F., Takahashi, B., Inoue, T., Fujiwara, S., Takagi, M.,

Iwamoto, T. & Kit, Y. (1998). Crystallographic studies on a family B DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J. Biochem.* **123**, 983–986.

28. Sakaguchi, N., Kuramitsu, S., Sakamoto, K., Shigesada, K., Fukushima, A., Watanabe, M., Arakawa, H. & Suzuki, K. (1990). Structural analysis of recombination with slanging plasmid DNA substrates and its involvement with homologous and illegitimate recombination systems in *Escherichia coli*. *Nucl. Acids Res.* **18**, 4819–4825.

29. Sakaguchi, N., Nakagawa, A., Ishibashi, S. & Suzuki, M. (1988). Magnesium-dependent exonuclease activities of the plasmid factory. *Proc. Natl Acad. Sci. USA*, **85**, 4826–4830.

30. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

31. Rossmann, M. G. (1990). Fast and accurate molecular replacement. *J. Appl. Cryst.* **23**, 426–433.

32. Navaza, J. (1994). Automatic molecular replacement. *Acta Crystallogr. sect. A*, **50**, 157–163.

33. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. sect. A*, **47**, 110–119.

34. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. sect. D*, **54**, 905–921.

35. Kleywegt, G. J. (1994). MAIN/LSQMAN: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 798–799.

36. Merritt, E. A. & Murphy, M. E. (1994). Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. sect. D*, **50**, 869–873.

37. Merritt, E. A. & Bacon, D. J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.

38. Nicholls, A., Sharp, K. A. & Honig, B. (1991). GRASP: Graphical Representation and Analysis of Surface Properties. *Biophys. Chem.* **39**, 281–291.

Edited by R. Huber

(Received 21 August 2000; received in revised form 8 December 2000; accepted 11 December 2000)

Crystal structure of an archaebacterial DNA polymerase

Yuxiang Zhao[1†], David Jeruzalmi[1†], Ismail Moarefi[1,2], Lorie Leighton[1,2], Roger Lasken[3] and John Kuriyan[1,2]*

Background: Members of the Pol II family of DNA polymerases are synonymous for chromosome replication in eukaryotes, and carry out highly processive DNA replication when associated with single-stranded processivity clamps. The sequences of the B polymerases are distinct from those of members of the well-studied Pol I family of DNA polymerases. The DNA polymerase from the archaebacterium *Sulfolobus solfataricus* (Sso Dp2 Pol II) is a member of the Pol II family that retains catalytic activity in elevated temperatures.

Results: The crystal structure of Sso Pol has been determined to 2.4 Å resolution. The architecture of Sso Pol B (type B) DNA polymerase resembles that of the DNA polymerase from the bacteriophage RB69, with which it shares only about 20% sequence identity. As in RB69, the central catalytic region of the DNA polymerase is flanked within the 'palm' subdomain and is strikingly similar in structure to the corresponding regions of Pol I type DNA polymerases. The structural scaffold that surrounds the catalytic core in D. *Tok* Pol is unrelated in structure to that of Pol I type polymerases. The 'thumb' protruding from the thumb domain of D. *Tok* Pol resembles the corresponding domains of HIV-1 RT and Pol I type DNA polymerases. The exonuclease domain in D. *Tok* Pol is located at the same position relative to the polymerase domain as seen in RB69, and on the opposite side of the palm subdomain compared to its location in Pol I type polymerases. The Pol II exonuclease domain of D. *Tok* Pol has structural similarity to DNA-binding domains. Sequence alignments suggest that this domain is conserved in the eukaryotic DNA polymerases δ and ε.

Conclusions: The structure of D. *Tok* Pol confirms that the mutation of homologs of the template acid activation of newly synthesized duplex DNA are likely to be distinct between Pol II and Pol I type DNA polymerases. However, the mechanism by which the newly synthesized product is kept out of the polymerizing active-site cleft seems to be quite different. The discovery of a domain that seems to be an RNA-binding module raises the possibility that Pol II family members interact with RNA.

Addresses: [1]Laboratories of Molecular Biophysics, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA and [2]Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA and [3]Molecular Genetics, Inc., 65 Valley Stream Parkway, Malvern, PA 19355, USA.

*The two first authors contributed equally to this work.

†Corresponding author. E-mail: kuriyan@rockefeller.edu

Key words: archaebacterium, *Sso*, DNA polymerase, *Tok*, family, RNA-binding domain, thermostability, X-ray crystallography

Received: 12 April 1999
Revisions requested: 27 May 1999
Revisions received: 9 June 1999
Accepted: 9 June 1999

Published: 28 September 1999

Structure October 1999, 7:1169–1185
http://structure.com/cgi/content/abstract/7/10/1169

©1999 Elsevier Science Ltd. All rights reserved.

Introduction

DNA polymerases can be classified into at least three families on the basis of sequence similarities to the three distinct DNA polymerases of *Escherichia coli*, Pol I, Pol II and Pol III [1–3]. Members of the Pol I family have been studied extensively, resulting in a comprehensive understanding of their functional properties and their structure [4–8]. In contrast to the detailed knowledge that is now available for the Pol I family, the Pol II and Pol III polymerases are poorly understood. The first crystal structure determined for a Pol II family member was that of the DNA polymerase of the bacteriophage RB69 (RB69 Pol) [9] and no structural information is currently available for any member of the Pol III family. Members of the Pol II (also known as Pol B or Pol α) and Pol III families carry out processive replication of chromosomal DNA during cell growth [3], and there is interest in further extending our knowledge of their structures and mechanisms. Sequence based DNA polymerases and the eukaryotic DNA polymerases δ, ε and ζ are members of the Pol II family [1].

The structure of RB69 Pol revealed that the general architecture of the core of the Pol II polymerases is strikingly similar to that of the Pol I polymerases [9]. Pol I polymerases are connected from three notable subdomains around the thumb, palm and fingers regions by analogy to α-helices first noted in the structure of the Klenow fragment of *E. coli* Pol I polymerase I [10]. In addition, Pol I DNA polymerases have a nucleotide-binding 3′–5′ exonuclease domain located before the thumb subdomain, near the region where duplex DNA exits the polymerase active site [4,5]. Besides the sequence conservation in catalytic domains, there is no significant sequence similarity between the polymerase domains of members of the Pol I and Pol II families [11]

However, the subdomain architecture of the Pol I family enzymes is the RB69 structure, even though the detailed structures of the palm domains are quite divergent [7]. The exonuclease domains of Pol I and Pol II DNA polymerases are distinct related in sequence and, not surprisingly, the structure of the exonuclease domain of RB69 resembles that of the Pol I type polymerases. Given the general similarity in the polymerase domains of the Pol I polymerases and RB69, the location of the exonuclease domain in RB69 was a surprise. In RB69 the 3'–5' exonuclease domain is located above the fingers and opposite the thumb subdomain, suggesting that the shuttling of DNA between the polymerization and proofreading sites must occur by a different mechanism to the Pol I DNA polymerases [7].

The mechanism of the Pol I family DNA polymerases is also undisputed (a detail [8]; Nature 28). The elements of nucleotide addition is mediated by two separate activities. These are located in the palm subdomain, at the base of a deep cleft in the polymerase domain. High-resolution crystal structures of the Pol I type DNA polymerases of T7 bacteriophage (T7 Pol) and *Thermus aquaticus* (Taq Pol) complexed to primer-template DNA and incoming nucleotide have been determined, revealing the mechanisms of nucleotide incorporation and selectivity to be visualized [9,11,28]. Although corresponding structural information for the Pol II family DNA polymerases is lacking, similarities in general organization of the polymerase core as well as conserved cross-motifs within crucial elements of the ternary palm subdomain suggest that genome fusion of the recognition of DNA will be similar in Pol II polymerases.

The DNA polymerase from the archaebacterium *Sulfolobus solfataricus* strain Tok 1B1 (Tok Pol) is a member of the Pol II family and has been characterizable DNA polymerase and 3'–5' exonuclease activities [13]. D. Tok Pol contains undetectable DNA polymerase activity after incubation at 95°C for one hour (81, unpublished results). The sequence of D. Tok Pol is very closely related to 70% identity to that of other archaebacterial DNA polymerases, such as those from *Pyrococcus furiosus* [14] and *Thermococcus gorgonarius* [15]. D. Tok Pol is also related to eukaryotic DNA polymerases α, δ and ε [16], sequence identity over 30% residues of the DNA polymerase core for the human δ sequence [17]. The archaebacterial genome also contains genes coding for proteins with close homology to proliferating cell nuclear antigen (PCNA), the DNA polymerase clamp in eukaryotes, as well as subunits of the clamp-loader complex, RFC (Replication factor C). It is likely that archaebacterial DNA polymerases achieve processivity by interaction with the ring-shaped PCNA sliding clamp, although direct evidence for such a mechanism is lacking.

We have determined the structure of D. Tok Pol at 2.6 Å resolution. D. Tok Pol shares less than 30% sequence identity with RB69 Pol, but the structures of the two enzymes' catalytic cores are closely. The structure reported here has been determined in the absence of DNA. Nevertheless, the close structural correspondence between the active sites of Pol I and Pol II DNA polymerases allows inferences to be made about the mode of DNA recognition by D. Tok Pol. The very N-terminal region of D. Tok Pol possesses a domain (residues 1–121) that is distantly related in structure to single-stranded DNA-binding domains (SSBDs) also known as RNA-recognition modules (RRMs) [18]. The structure of the 3'–5' proofreading exonuclease domain of D. Tok Pol is similar to those of the Pol I type polymerases. However, its location, relative to the palm subdomain, resembles the location seen in RB69 Pol rather than the Pol I type polymerases [6,10,11]. The structure of D. Tok Pol reported here provides further evidence that the mode of DNA-template recognition and the detailed editing chemical established for the Pol II family by the structure of RB69 Pol is valid for the entire Pol II family.

Results and discussion

Structure determination

Crystals of D. Tok Pol have been obtained from 2-hydroxyethyldisulfide (HPED) (Sigma-D) and polyethylene glycol (PEG) 400 (Merck-H). Both crystal forms are orthorhombic ($P2_12_12_1$, $a = 59.8$ Å, $b = 189.8$ Å, $c = 163.2$ Å for Native I and $a = 60.1$ Å, $b = 190.6$ Å, $c = 155.9$ Å for Native II). Experimental phases (7.6Å) to 4 Å were obtained from four heavy-atom derivatives by multiple isomorphous replacement and anomalous scattering (MIRAS) [19]. Phases were improved via iterative cycles of real-space density modification, averaging of solvent flipping and negative density truncation, using SOLOMON [20,21]. The resulting electron-density map allowed the chain to be traced easily. Especially, with easily determination of sequence register. The model was refined to 2.6 Å against data for Native II (R value = 24.2%, R_{free} = 29.6%) and subsequently, to 2.0 Å against data for Native I (R value = 25.3%, R_{free} = 29.9%), using CNS [22]. The model for Native II is somewhat more complete than the Materials are available in this and is used for most of the discussion. This model includes 790 residues from 1 to 776 in Native II. Arginine 450–456 and 655–676 are not visible in the electron-density map and are not included in the model.

General description of the structure

D. Tok Pol (Figure 1) is composed of a polymerase domain (residues 396–775) and an exonuclease domain (residues 122–385), as well as an N-terminal domain (residues 1–121) that is not found in Pol I type DNA polymerases [4]. The polymerase domain is further comprised of three smaller subdomains, termed the thumb (residues 669–776), palm (residues 560–643 and 500–619) and fingers (residues 480–499). The structures of the ΔPED and PED400 crystal forms of D. Tok Pol are very similar

Table 1

Data collection, structure determination and refinement statistics.

	Resolution (Å)	Number of unique reflections	Completeness (%)	R_{sym} (%)	R_{iso} (%)	Sites	Phasing power*	Figures of merit
Native data	–	–	–	–	–	–	–	–
Native 9	30.0–2.6	12,009	92.9(53.8)	5.2(5.6)	–	–	–	–
Native 2	30.0–2.7	10,949	90.2(50.8)	4.0(1.8)	–	–	–	–
Mercury analysis	–	–	–	–	–	4	0.267	
Hg	500–8.0	10,284	91.5(93.3)	8.4(3.9)	19.1	6	1.34(0.89)	0.219
Hg	30.0–3.0	14,963	94.3(79.4)	9.5(5.6)	15.5	7	1.15(0.68)	0.195
Hg	30.0–3.2	10,421	89.9(57.4)	10.2(7.5)	19.3	6	1.69(0.95)	0.227
Heavy-atom refinement	Number of Reflections ($|F| > 2\sigma$)			$R_{cryst} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$ (%)	Final number of atoms	Rms deviations from ideal	Rms/Min angles (°)	Rms/Min B-values (Å²)
Native9	500–2.6	51,691		24.3(29.0	2,137	0.018/2.73	2.81/331	3.49/–
Native2	1000–2.8	37,278		25.2(31.8	2,145	0.020/2.73	1.30/4.79	1.41/9

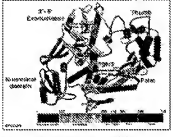
$R_{sym}(I) = 100 \Sigma |(I) - \langle I \rangle| / \Sigma(I)$, where I is the observed intensity and $\langle I \rangle$ is the mean intensity obtained from multiple observations of symmetry-related reflections after rejection of outliers. *Phasing power = $\Sigma |F_{calc}| / \Sigma ||F_{calc}| \pm |F_{obs}||$, where $|F_{calc}|$ and $|F_{obs}|$ are the absolute and isomorphous scattering amplitudes of the heavy-atom derivative. †Figure of merit = $|\Sigma P(\alpha) e^{i\alpha} / \Sigma P(\alpha)|$, where α is the phase and $P(\alpha)$ is the probability that α is correct. ‡Native 9 / Native 2 / Hg1 / Hg2 / Hg3. §$R_{cryst} = \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$ for all data between 10.0 and 2.6 Å. ¶Root-mean-square deviations from ideal bond lengths (Å) and angles (°). #Overall / main-chain / side-chain B-values (Å²).

in terms of the individual subunits. The major difference between the two structures is a rotation of ~5–10° in the orientation of the exonuclease domain with respect to the thumb subdomain.

The domains of D. Tth Pol are arranged as an irregularly shaped flattened ring with a central cavity formed near the polymerase active site. The mostly α-helical thumb subdomain forms one side of the active-site cleft and makes contacts with the exonuclease domain (Figure 1). The structures of the thumb domains of various polymerases are often correlated in structure. However, in all cases where structures are available the thumb domain is seen to exhibit an important role by flanking contacts with duplex DNA as it exits the polymerase active site [4]. The D. Tth Pol structure has been determined in the absence of DNA, and a portion of the thumb subdomain that is likely to contact DNA structures (65–103) is disordered. This is commonly observed for the corresponding regions of other polymerases in the absence of substrate [9,11–14]. In the DNA polymerases from bacteriophage T4 and RB69, the thumb subdomains also provide a C-terminal sliding-clamp that interacts with the processivity factor [15,16]. In D. Tth Pol, the corresponding region (residues 757–773) is disordered.

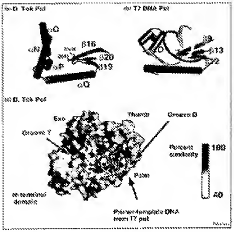
The second region of the active-site cleft is occupied by the gene subdomain and includes residues important for substrate discrimination and the catalysis of the polymerase reaction. In D. Tth Pol, the palm is organized around seven β strands (β1a, β1b, β2b) flanked by six α helix (αO) (Figures 1 & 3a). It contains two disulfide bonds (Cys428–Cys442, Cys505–Cys519) that have not been previously observed in palm subdomains and which may be important for thermostability (Figure 3).

Figure 3



Structure of D. Tth Pol. The structure is represented by polypeptide ribbons, strands, and coils; each is colour coded according to its function. Two dyad-symmetric molecules of DNA are bound to the exonuclease domain. The active site of the exonuclease domain is located ~30 Å from the polymerase active site. Both DNA molecules are shown in stick representation. Regions of the polypeptide chain that could not be modeled in the present structure because of disorder are indicated by dotted lines. The various domains and subdomains and their boundaries are indicated in the key.

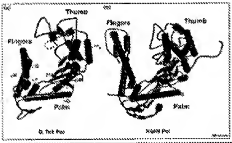
Figure 2



Comparison of DNA polymerase structures. [illegible]

The central elements of the palm subdomains from polymerases belonging to the Pol I and Pol II families can be aligned closely (the root mean square deviation (rmsd) in Cα positions for strands β18, β19, β20 and helix αQ is in the range of 0.6–2.0 Å), indicating a structural conservation of function. There are two residues in the palm domains

Figure 3



Comparison of the structures of D. Tok Pol and RB69 Pol. The structures of (a) D. Tok Pol and (b) RB69 Pol are [illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4603542v1

of Pol I polymerases that are crucial for polymerase activity, because they coordinate two metal ions [2,19,21,27]. The corresponding residues in D. Tok Pol are Asp406 and Asp543 (Figure 2). No metal ions are, however, visible in our electron-density maps.

The fingers subdomain in D. Tok Pol consists of a set of antiparallel α helices (αC, αD, αE, Figure 2). These helices are shorter in length than the corresponding elements of RB69 Pol and a helical segment that connects helices G and N in RB69 Pol is missing altogether (Figure 3). The fingers domain of D. Tok Pol is anchored to external structures to that of Pol I type polymerases (Figure 2). However, helix αE in D. Tok Pol is positioned distantly to helix O in Pol I polymerases (Figure 3), and is likely to play an analogous and essential role in recognition of the incoming nucleotide [9–11,28].

The PolB exonuclease domain in D. Tok Pol is located opposite the thumb subdomain and above the fingers subdomain, as noted for RB69 Pol. It contains two metal ions – presumably Mg^{2+} ligated to Asp411 and Glu141 (Figure 3). The position of this domain relative to the polymerase active site is distinct from the arrangement in a Pol I type polymerase. The conservation between RB69 and D. Tok Pol of the location of the exonuclease domain suggests that this is a characteristic feature of the B type polymerases. The presence of the D. Tok Pol 3′–5′ exonucleolytic domain resembles those associated with other DNA polymerases [29,30]. The 3′–5′ exonuclease domains from the Pol I (*E. coli*, *T. aquaticus*, *Bacillus subtilis*, bacteriophage T7) or Pol B (RB69) polymerase families can be aligned with each other closely (small to medium-size elements [31], [32], [34] and helices αB and αI in the range of 1.0–2.8 Å). This alignment suggests a precise molecular environment with universal binding, catalytic and metal binding in a conformery montage (Figure 4) [6].

The arrangement of the N-terminal, exonuclease, and polymerase domains creates two deep grooves leading into and out of the polymerase active site. The D groove clasps duplex-DNA binding, following the nonconformer of [7]; a 4-strand antiparallel β sheet extends from the thumb subdomain and encircles a region of positive electrostatic potential. The T groove, the complete DNA binding track away from the active site in the opposite direction and is located below the fingers subdomain. A small β-sheet tube relating channel leads from the polymerase domain to the exonuclease active site (Figure 2c).

We have used the structure of T7 Pol bound at primer-template DNA to model DNA into D. Tok Pol (Figure 2a). Superposition of the polymerase subdomains of the two polymerases shows that remarkably few bad contacts are formed between the DNA from T7 Pol and amino in the D. Tok Pol model. The two regions that does collide

Figure 3
Structure-based sequence alignment of archaeactional DNA polymerases. Residues 1–1000 of D. Tok Pol (GenBank accession number AF 182814) and RB69 DNA polymerase (GenBank accession number U27-34, U35-136) are aligned. Residues 1–280, 280–340, and 367–513 which comprise the exonuclease, fingers, and thumb subdomains of D. Tok Pol are aligned to RB69. Secondary structure elements are shown above the sequence for D. Tok Pol and below the sequence for RB69. Helices are labeled with capital letters and β strands with numbers. Conserved sequence motifs are shown in green (Exo) and dark blue (Pol). Conserved sequence elements and motifs in the catalytic core are shown in green (Exo) and dark blue (Pol). Conserved residues are shown in white letters on a red background. Residues that are conserved in all but one sequence are shown in red letters. Residues that are conserved in all but two sequences are shown in blue letters. Residues that are conserved in all but three sequences are shown in green letters. Residues that are conserved in all but four sequences are shown in black letters. Residues that are conserved in all but five sequences are shown in gray letters. Residues that are conserved in all but six sequences are shown in light gray letters. Residues that are conserved in all but seven sequences are shown in light gray letters. Residues that are conserved in all but eight sequences are shown in light gray letters. Residues that are conserved in all but nine sequences are shown in light gray letters. Residues that are conserved in all but ten sequences are shown in light gray letters. Residues that are conserved in all but eleven sequences are shown in light gray letters. Residues that are conserved in all but twelve sequences are shown in light gray letters. Residues that are conserved in all but thirteen sequences are shown in light gray letters. Residues that are conserved in all but fourteen sequences are shown in light gray letters. Residues that are conserved in all but fifteen sequences are shown in light gray letters. Residues that are conserved in all but sixteen sequences are shown in light gray letters. Residues that are conserved in all but seventeen sequences are shown in light gray letters. Residues that are conserved in all but eighteen sequences are shown in light gray letters. Residues that are conserved in all but nineteen sequences are shown in light gray letters. Residues that are conserved in all but twenty sequences are shown in light gray letters. Residues that are conserved in all but twenty-one sequences are shown in light gray letters. Residues that are conserved in all but twenty-two sequences are shown in light gray letters. Residues that are conserved in all but twenty-three sequences are shown in light gray letters. Residues that are conserved in all but twenty-four sequences are shown in light gray letters. Residues that are conserved in all but twenty-five sequences are shown in light gray letters. Residues that are conserved in all but twenty-six sequences are shown in light gray letters. Residues that are conserved in all but twenty-seven sequences are shown in light gray letters. Residues that are conserved in all but twenty-eight sequences are shown in light gray letters. Residues that are conserved in all but twenty-nine sequences are shown in light gray letters. Residues that are conserved in all but thirty sequences are shown in light gray letters. Residues that are conserved in all but thirty-one sequences are shown in light gray letters. Residues that are conserved in all but thirty-two sequences are shown in light gray letters. Residues that are conserved in all but thirty-three sequences are shown in light gray letters. Residues that are conserved in all but thirty-four sequences are shown in light gray letters. Residues that are conserved in all but thirty-five sequences are shown in light gray letters. Residues that are conserved in all but thirty-six sequences are shown in light gray letters. Residues that are conserved in all but thirty-seven sequences are shown in light gray letters. Residues that are conserved in all but thirty-eight sequences are shown in light gray letters. Residues that are conserved in all but thirty-nine sequences are shown in light gray letters. Residues that are conserved in all but forty sequences are shown in light gray letters. Residues that are conserved in all but forty-one sequences are shown in light gray letters. Residues that are conserved in all but forty-two sequences are shown in light gray letters. Residues that are conserved in all but forty-three sequences are shown in light gray letters. Residues that are conserved in all but forty-four sequences are shown in light gray letters. Residues that are conserved in all but forty-five sequences are shown in light gray letters. Residues that are conserved in all but forty-six sequences are shown in light gray letters. Residues that are conserved in all but forty-seven sequences are shown in light gray letters. Residues that are conserved in all but forty-eight sequences are shown in light gray letters. Residues that are conserved in all but forty-nine sequences are shown in light gray letters. Residues that are conserved in all but fifty sequences are shown in light gray letters. Residues that are conserved in all but fifty-one sequences are shown in light gray letters. Residues that are conserved in all but fifty-two sequences are shown in light gray letters. Residues that are conserved in all but fifty-three sequences are shown in light gray letters. Residues that are conserved in all but fifty-four sequences are shown in light gray letters. Residues that are conserved in all but fifty-five sequences are shown in light gray letters. Residues that are conserved in all but fifty-six sequences are shown in light gray letters. Residues that are conserved in all but fifty-seven sequences are shown in light gray letters. Residues that are conserved in all but fifty-eight sequences are shown in light gray letters. Residues that are conserved in all but fifty-nine sequences are shown in light gray letters. Residues that are conserved in all but sixty sequences are shown in light gray letters. Residues that are conserved in all but sixty-one sequences are shown in light gray letters. Residues that are conserved in all but sixty-two sequences are shown in light gray letters. Residues that are conserved in all but sixty-three sequences are shown in light gray letters. Residues that are conserved in all but sixty-four sequences are shown in light gray letters. Residues that are conserved in all but sixty-five sequences are shown in light gray letters. Residues that are conserved in all but sixty-six sequences are shown in light gray letters. Residues that are conserved in all but sixty-seven sequences are shown in light gray letters. Residues that are conserved in all but sixty-eight sequences are shown in light gray letters. Residues that are conserved in all but sixty-nine sequences are shown in light gray letters. Residues that are conserved in all but seventy sequences are shown in light gray letters. Residues that are conserved in all but seventy-one sequences are shown in light gray letters. Residues that are conserved in all but seventy-two sequences are shown in light gray letters. Residues that are conserved in all but seventy-three sequences are shown in light gray letters. Residues that are conserved in all but seventy-four sequences are shown in light gray letters. Residues that are conserved in all but seventy-five sequences are shown in light gray letters. Residues that are conserved in all but seventy-six sequences are shown in light gray letters. Residues that are conserved in all but seventy-seven sequences are shown in light gray letters. Residues that are conserved in all but seventy-eight sequences are shown in light gray letters. Residues that are conserved in all but seventy-nine sequences are shown in light gray letters. Residues that are conserved in all but eighty sequences are shown in light gray letters. Residues that are conserved in all but eighty-one sequences are shown in light gray letters. Residues that are conserved in all but eighty-two sequences are shown in light gray letters. Residues that are conserved in all but eighty-three sequences are shown in light gray letters. Residues that are conserved in all but eighty-four sequences are shown in light gray letters. Residues that are conserved in all but eighty-five sequences are shown in light gray letters. Residues that are conserved in all but eighty-six sequences are shown in light gray letters. Residues that are conserved in all but eighty-seven sequences are shown in light gray letters. Residues that are conserved in all but eighty-eight sequences are shown in light gray letters. Residues that are conserved in all but eighty-nine sequences are shown in light gray letters. Residues that are conserved in all but ninety sequences are shown in light gray letters. Residues that are conserved in all but ninety-one sequences are shown in light gray letters. Residues that are conserved in all but ninety-two sequences are shown in light gray letters. Residues that are conserved in all but ninety-three sequences are shown in light gray letters. Residues that are conserved in all but ninety-four sequences are shown in light gray letters. Residues that are conserved in all but ninety-five sequences are shown in light gray letters. Residues that are conserved in all but ninety-six sequences are shown in light gray letters. Residues that are conserved in all but ninety-seven sequences are shown in light gray letters. Residues that are conserved in all but ninety-eight sequences are shown in light gray letters. Residues that are conserved in all but ninety-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred sequences are shown in light gray letters. Residues that are conserved in all but one hundred one sequences are shown in light gray letters. Residues that are conserved in all but one hundred two sequences are shown in light gray letters. Residues that are conserved in all but one hundred three sequences are shown in light gray letters. Residues that are conserved in all but one hundred four sequences are shown in light gray letters. Residues that are conserved in all but one hundred five sequences are shown in light gray letters. Residues that are conserved in all but one hundred six sequences are shown in light gray letters. Residues that are conserved in all but one hundred seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred ten sequences are shown in light gray letters. Residues that are conserved in all but one hundred eleven sequences are shown in light gray letters. Residues that are conserved in all but one hundred twelve sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirteen sequences are shown in light gray letters. Residues that are conserved in all but one hundred fourteen sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifteen sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixteen sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventeen sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighteen sequences are shown in light gray letters. Residues that are conserved in all but one hundred nineteen sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred twenty-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred thirty-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred forty-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred fifty-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred sixty-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred seventy-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred eighty-nine sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-one sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-two sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-three sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-four sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-five sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-six sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-seven sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-eight sequences are shown in light gray letters. Residues that are conserved in all but one hundred ninety-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred sequences are shown in light gray letters. Residues that are conserved in all but two hundred one sequences are shown in light gray letters. Residues that are conserved in all but two hundred two sequences are shown in light gray letters. Residues that are conserved in all but two hundred three sequences are shown in light gray letters. Residues that are conserved in all but two hundred four sequences are shown in light gray letters. Residues that are conserved in all but two hundred five sequences are shown in light gray letters. Residues that are conserved in all but two hundred six sequences are shown in light gray letters. Residues that are conserved in all but two hundred seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred ten sequences are shown in light gray letters. Residues that are conserved in all but two hundred eleven sequences are shown in light gray letters. Residues that are conserved in all but two hundred twelve sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirteen sequences are shown in light gray letters. Residues that are conserved in all but two hundred fourteen sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifteen sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixteen sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventeen sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighteen sequences are shown in light gray letters. Residues that are conserved in all but two hundred nineteen sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred twenty-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred thirty-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred forty-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred fifty-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred sixty-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred seventy-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred eighty-nine sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-one sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-two sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-three sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-four sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-five sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-six sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-seven sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-eight sequences are shown in light gray letters. Residues that are conserved in all but two hundred ninety-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred sequences are shown in light gray letters. Residues that are conserved in all but three hundred one sequences are shown in light gray letters. Residues that are conserved in all but three hundred two sequences are shown in light gray letters. Residues that are conserved in all but three hundred three sequences are shown in light gray letters. Residues that are conserved in all but three hundred four sequences are shown in light gray letters. Residues that are conserved in all but three hundred five sequences are shown in light gray letters. Residues that are conserved in all but three hundred six sequences are shown in light gray letters. Residues that are conserved in all but three hundred seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred ten sequences are shown in light gray letters. Residues that are conserved in all but three hundred eleven sequences are shown in light gray letters. Residues that are conserved in all but three hundred twelve sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirteen sequences are shown in light gray letters. Residues that are conserved in all but three hundred fourteen sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifteen sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixteen sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventeen sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighteen sequences are shown in light gray letters. Residues that are conserved in all but three hundred nineteen sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred twenty-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred thirty-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred forty-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred fifty-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred sixty-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred seventy-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred eighty-nine sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-one sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-two sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-three sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-four sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-five sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-six sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-seven sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-eight sequences are shown in light gray letters. Residues that are conserved in all but three hundred ninety-nine sequences are shown in light gray letters. Residues that are conserved in all but four hundred sequences are shown in light gray letters. Residues that are conserved in all but four hundred one sequences are shown in light gray letters. Residues that are conserved in all but four hundred two sequences are shown in light gray letters. Residues that are conserved in all but four hundred three sequences are shown in light gray letters. Residues that are conserved in all but four hundred four sequences are shown in light gray letters. Residues that are conserved in all but four hundred five sequences are shown in light gray letters. Residues that are conserved in all but four hundred six sequences are shown in light gray letters. Residues that are conserved in all but four hundred seven sequences are shown in light gray letters. Residues that are conserved in all but four hundred eight sequences are shown in light gray letters. Residues that are conserved in all but four hundred nine sequences are shown in light gray letters. Residues that are conserved in all but four hundred ten sequences are shown in light gray letters. Residues that are conserved in all but four hundred eleven sequences are shown in light gray letters. Residues that are conserved in all but four hundred twelve sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirteen sequences are shown in light gray letters. Residues that are conserved in all but four hundred fourteen sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifteen sequences are shown in light gray letters. Residues that are conserved in all but four hundred sixteen sequences are shown in light gray letters. Residues that are conserved in all but four hundred seventeen sequences are shown in light gray letters. Residues that are conserved in all but four hundred eighteen sequences are shown in light gray letters. Residues that are conserved in all but four hundred nineteen sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-one sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-two sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-three sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-four sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-five sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-six sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-seven sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-eight sequences are shown in light gray letters. Residues that are conserved in all but four hundred twenty-nine sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-one sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-two sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-three sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-four sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-five sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-six sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-seven sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-eight sequences are shown in light gray letters. Residues that are conserved in all but four hundred thirty-nine sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-one sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-two sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-three sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-four sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-five sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-six sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-seven sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-eight sequences are shown in light gray letters. Residues that are conserved in all but four hundred forty-nine sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-one sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-two sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-three sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-four sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-five sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-six sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-seven sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-eight sequences are shown in light gray letters. Residues that are conserved in all but four hundred fifty-nine sequences are shown in light gray letters. Residues that are conserved in all but four hundred sixty sequences are shown in light gray letters. Residues that are conserved in all but four hundred sixty-one sequences are shown in light gray letters. Residues that are conserved in all but four hundred sixty-

Figure 8

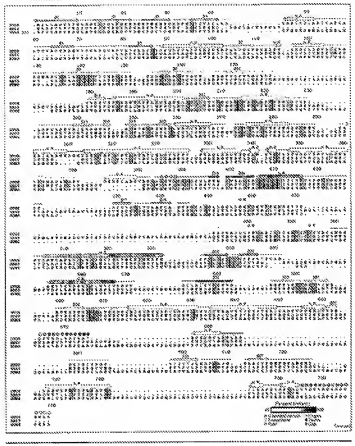


Figure 5

Structure-based sequence alignment of D. Tok Pol (D.TOK), RB69 Pol (RB69) and human Pol δ (HUMD). [illegible]

closely (Figure 3). Not surprisingly, the regions of highest sequence similarity are concentrated in and around the exonuclease and polymerase active sites (Figures 2c,5). Despite the low overall sequence identity, the individual subdomains in the two structures superimpose well (rms most in Cα positions in the fingers, thumb and palm subdomains is in the range of 0.8 to 1.5 Å). Moreover, the overall arrangement of domains and subdomains with respect to each other is preserved in the two polymerases, strengthening the proposal that Pol B DNA polymerases share a common architecture (Figure 3).

One difference between the overall structures of D. Tok Pol and RB69 Pol concerns the orientation of the exonuclease domain with respect to the rest of the structure. When the two polymerases are superimposed on their respective palm subdomains it is seen that the exonuclease domain of RB69 is rotated inwards by ~8°, burying the active site in a solvent-inaccessible configuration [7]. In contrast, the exonuclease domain in D. Tok Pol has its active site essentially exposed to solvent. It is possible that conformational changes between open and closed configurations of the exonuclease domain are a part of the functional cycle of the protein, particularly as the two different forms of D. Tok Pol differ in the orientation of the exonuclease domain (not shown).

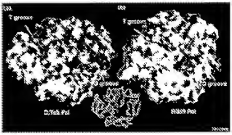
One interesting difference between D. Tok Pol and RB69 Pol is that the former is a thermostable DNA polymerase whereas the latter is not. Unfortunately, attempts to identify features in the D. Tok Pol structure that might be correlated with thermostability is complicated by the very low sequence similarity between the two enzymes. One feature that does stand out, however, is the increased formation of arrays of ionic interactions on the surface of D. Tok Pol when compared to that of RB69 Pol (Figure 6). The formation of networks of ionic interactions has been found to correlate with thermostability in other proteins [16,32,33].

Generally, D. Tok Pol subdomains tend to be more compact, with smaller helices and shorter loops than are found in RB69 Pol, a feature that may be another important source of thermostability. For example, the palm subdomain displays close structural conservation of elements near the catalytic aspartate residues. However, helix αR in D. Tok Pol is much shorter than its counterpart in RB69 Pol, and a small subdomain in front of the palm subdomain is entirely missing in D. Tok Pol (Figures 3,5). Deletion of these elements is also seen in a representative set of archaebacterial DNA polymerases [13,34]. Likewise, the fingers subdomain is missing a large insertion from its tip in D. Tok Pol (Figures 3,5). However, the RB69 fingers insertion most probably plays a T4 phage-specific role, as it is also missing from the alignments of archaebacterial DNA polymerases and eukaryotic polymerases δ (Figure 5).

The N-terminal domain resembles RNA-binding domains

The N-terminal domain of D. Tok Pol has no corresponding element in Pol B type polymerases. Analysis of the

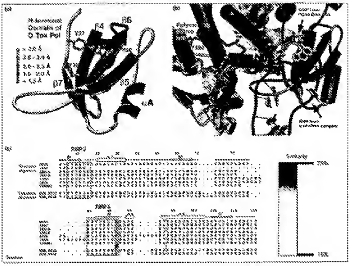
Figure 6



Comparison of surface charges in D. Tok Pol and RB69 Pol. Electrostatic surface representations of (a) D. Tok Pol and (b) RB69 Pol in the same orientation after superposition of their palm subdomains. [illegible] D. Tok Pol has a striking [illegible] not seen in RB69 Pol. A representation of D. Tok Pol [illegible] is included for orientation.

USSN 10/089,027
Atty. Docket No. 2007651-0001

4603542v1



and bind to single-stranded RNA. Two conserved sequence motifs, referred to as RNP1 (ribonucleoprotein 1) and RNP2, provide aromatic and charged residues that are important for RNA recognition [55] (Figure 7).

The N-terminal domain of D. Tok Pol can be superimposed closely onto the core secondary structural elements of RBDs from the U1A spliceosomal protein [53], ribosomal protein S6 [56], the heterogeneous ribonucleoprotein (hnRNP) proteins (two RBD domains) [52,54] and the anticodon-binding domain from *T. thermophilus* phenylalanyl-tRNA synthetase [57]. The rmsds in Cα positions for these superpositions are in the range of 1.5–2.0 Å (Figure 7c). Differences between the structures of the loops in the N-terminal domain of D. Tok Pol and those of the RNA-binding domains are within the range of structural variation seen in the various RNA-binding domains.

There is no evidence at present to suggest that the N-terminal domain of D. Tok Pol binds RNA. However, comparison with the structures of RNA complexes of RNA-binding domains shows that the N-terminal domain might in fact be a functional RNA-binding domain (Figure 7). In particular, three aromatic residues in the N-terminal domain (Tyr37, Tyr39 and Tyr88) could interact with RNA bases in a manner similar to that seen in crystal structures of RNA bound to RNA-binding domains [53] (Figure 7). Interestingly, these residues are located near the position of a guanosine triphosphate molecule that is found bound to the N-terminal domain of RB69 Pol [1] (Figure 7b). The DNA polymerases from bacteriophage T4 and its distant relative bacteriophage RB69 bind specifically to the ribosome-binding site of their own mRNA (messenger RNA), repressing its translation [40–42]. The N-terminal domains of T4 Pol and RB69 Pol are smaller than that of D. Tok Pol. In the RB69 Pol structure, the N-terminal domain seems to form an 'incomplete' RNA-binding domain (Figure 7c).

There is no significant overall sequence similarity between the N-terminal domain of D. Tok Pol and RNA-binding domains, which is why the presence of this fold was not recognized previously (Figure 7c). Comparison of the sequences of other archaebacterial DNA polymerases and human polymerases δ and ε suggests that a corresponding structural element is likely to be found in these polymerases as well (Figure 7c). The sequence alignment in this region is unambiguous for the archaebacterial DNA polymerases. For eukaryotic polymerases the alignment is less certain, but it seems to conserve the essential aromatic character of the RNP motifs (Figure 7c). Confirmation of the presence of these domains along with their ability to bind RNA and their precise role in eukaryotic DNA synthesis awaits future structural and functional studies.

Biological implications

The structure of the DNA polymerase from the archaebacterium *Desulfurococcus* strain Tok, D. Tok Pol, reveals a strong similarity to the DNA polymerase from bacteriophage RB69. It also reveals the presence of an N-terminal domain that has structural similarity to RNA-binding domains from the U1A spliceosomal protein, ribosomal protein S6, the hnRNP proteins and the anticodon-binding domain from *T. thermophilus* phenylalanyl-tRNA synthetase. Although the structure in the immediate vicinity of the central catalytic region of the polymerase domain closely resembles that of Pol I type DNA polymerases, the overall architecture of D. Tok Pol and the placement of the exonuclease domain is strikingly different. The similarity between D. Tok Pol and RB69 Pol suggests that these two structures are representative of a common Pol II polymerase fold. Members of this family carry out chromosomal DNA replication in eukaryotes, including humans, and yet there is no structural information available for any eukaryotic member of this family. While this manuscript was being prepared, the structure of another archaebacterial DNA polymerase, that from the organism *Thermococcus gorgonarius* has been reported [58]. The D. Tok Pol structure reported here, along with the RB69 Pol structure and the structure of the *Thermococcus gorgonarius* DNA polymerase, should now make it possible to generate reliable structural models for eukaryotic DNA polymerases.

Materials and methods

Protein expression and purification

The D. Tok Pol bacterial expression vector and [illegible] and sequence were generous gifts of [illegible] Technology Corporation. [illegible] D. Tok Pol [illegible] [illegible] D. Tok Pol [illegible] 80°C [illegible] 30 min [illegible] [illegible] [illegible] (pH = 7.5), 50 mM [illegible] for crystallization trials.

Crystallization, [illegible], and data collection

Crystals of D. Tok Pol [illegible] were prepared from 100 mM TRIS-HCl (pH = 8.5), [illegible] $MgCl_2$, 200 mM $(NH_4)_2SO_4$, [illegible] [illegible] 20°C. [illegible] [illegible] [illegible] [illegible] group $P2_12_12_1$, [illegible] a = [illegible] Å, b = [illegible] Å, c = [illegible] Å, [illegible] [illegible] [illegible] [illegible] [illegible]

Acknowledgements

References

51 Bendtsen JD, Kiemer L, Fausbøll A, Brunak S (2005) Non-classical protein secretion in bacteria. BMC Microbiol 5: 58

52 Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783–795

53 Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S (2005) Prediction of twin-arginine signal peptides. BMC Bioinformatics 6: 167

54 Bernstein HD, Poritz MA, Strub K, Hoben PJ, Brenner S, Walter P (1989) Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. Nature 340: 482–486

55 Bolhuis A, Tjalsma H, Smith HE, de Jong A, Meima R, Venema G, Bron S, van Dijl JM (1999) Evaluation of bioinformatic prediction rules for the membrane-bound subtilisin protease SipW from Bacillus subtilis. Microbiology 145: 2159–2167

56 Bolhuis A, Tjalsma H, Smith HE, Kontinen VP, Venema G, Bron S, van Dijl JM (1998) Functional analysis of the membrane-bound subtilisin protease SipW from Bacillus subtilis. J Biol Chem 273: 21217–21224

57 Bolhuis A, Venema G, Kontinen VP, Meima R, Tjalsma H, Bron S, van Dijl JM (2001) Functional analysis of the cytoplasmic serine protease domain of the prokaryotic preprotein translocase-associated protease. J Biol Chem 276: 1183–1190

58 Bolhuis H, Molenaar D, Poelarends GJ, Konings WN (1998) Characterization of the tetrahydrofolate biosynthesis operon of Lactobacillus plantarum. J Bacteriol 180: 4124–4127

59 Bolhuis H, Poelarends GJ, van Veen HW, Poolman B, Driessen AJ, Konings WN (1996) Energetics and mechanism of drug transport mediated by the lactococcal multidrug transporter LmrP expressed in Lactococcus lactis. J Biol Chem 271: 24544–24548

60 Bolhuis H, van Veen HW, Poolman B, Driessen AJ, Konings WN (1997) Mechanisms of multidrug transporters. FEMS Microbiol Rev 21: 55–84

61 Bolhuis H, van Veen HW, Poolman B, Driessen AJ, Konings WN (1997) Multidrug transport in the spectrum of bacteria. FEMS Microbiol Lett 150: 195–201

62 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

63 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

64 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

65 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

66 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

67 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

68 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

69 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

70 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

71 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

72 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

73 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

74 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

75 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

76 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

77 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

78 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

79 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

80 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

81 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

82 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

83 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

84 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

85 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

86 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

87 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

88 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

89 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

90 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

91 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

92 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

93 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

94 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

95 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

96 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

97 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

98 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

99 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

100 Bolhuis H, Venema G, Kok J, van Veen HW, Konings WN (1995) Molecular characterization of the tetrahydrofolate biosynthesis operon of Lactococcus lactis. J Bacteriol 177: 1732–1738

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 3669–3674, March 1999
Biochemistry

Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*

Archaea/structure/disulfide bonds/replication/mechanism/exonuclease

KARL-PETER HOPFNER*†‡, ANDREAS EICHINGER*, RICHARD A. ENGH*§, FRANK LAUE¶, WALDEMAR ANKENBAUER¶, ROBERT HUBER*, AND BERNHARD ANGERER¶||

*Abteilung Strukturforschung, Max Planck-Institut für Biochemie, D-82152 Martinsried, Germany; and ¶Roche Diagnostics, D-82372 Penzberg, Germany

Contributed by Robert Huber, January 21, 1999

ABSTRACT Many known archaeal DNA polymerases belong to the pol type B family, which also includes the DNA replication polymerases of eukaryotes, but molecules with high fidelity in extreme conditions. We describe here the 2.3 Å resolution crystal structure of a DNA polymerase from the archaeon *Thermococcus gorgonarius* and identify structural features of the fold and the active site that are likely responsible for its thermostable function. Comparison with the monomeric B type DNA polymerase gp43 of the bacteriophage RB69 highlights thermophilic adaptations, which include the presence of two disulfide bonds and an enhanced electrostatic complementarity at the DNA-protein interface. In contrast to gp43, several loops at the exonuclease and thumb domains are more closely packed, this apparently blocks primer binding to the exonuclease active site. A physiological role of this "closed" conformation is unknown but may represent a polymerase mode, in contrast to an editing mode with an open exonuclease site. This archaeal B DNA polymerase structure provides a starting point for structure-based design of polymerases in fundamental applications in biotechnology and the development of antiviral or anticancer agents.

Propagation of cells requires faithful DNA replication. This is performed in vivo by DNA polymerases (pols), which select the appropriate dNTP at the nascent DNA primer strand to match its paired template. Different families of pols are encoded in different DNA polymerization processes including not only DNA replication (1, 2) but also repair and recombination (3, 4). A heterogeneity also manifested by varying polypeptide structures and subunit compositions (5, 6). Some pols strengthen polymerase activity with $3' \rightarrow 5'$ exonuclease activity (proofreading activity) within $5' \rightarrow 3'$ "exonuclease-specific endonuclease" activity often located in extensive structural elements on the same polypeptide chain (6–8).

Crystal structures are available for seven known polymerase families, including the A family DNA polymerases (9–14), pol β (15–17), HIV reverse transcriptase (18–20), and recently, the B family pol gp43 from bacteriophage RB69 (21). All show a functional polymerase structure, which resembles a right hand held by the palm, fingers and thumb domains (also ref. 7 for review). Although the fingers and thumb domains are highly divergent among the different families, the palm domains, which contain the conserved catalytic sequence residues, show a similar topology among all families, except pol β. The polymerase nucleotide transfer site studied in detail for the A family polymerases, HIV reverse transcriptase, and pol β, and was shown to involve two metal ions (summarized in ref. 7).

Considerably less is known for the family of type B pols, which use replicative enzymes in eukaryotes and some thermophilic archaea (22, 23). The structure of gp43 from bacteriophage RB69 (21) provided an excellent first insight into this family in addition to the three polymerase domains, gp43 contains an $3' \rightarrow 5'$ exonuclease domain and an N-terminal domain. The exonuclease and palm domains share the topology and active site of A family enzymes, implying similar metal-assisted mechanisms for polymerase and exonuclease activities (21). The thumb and finger domains are apparently specialized to the other polymerase families. The function of the N-terminal domain remains unknown, but may help to confer the unique processive replication apparatus (21).

Much is known about the replication of phages [24–26], viruses [1, 27], Protozoa, [28] and Eubacteria [1, 3, 29, 30], which in general involves pols that also process primers, RNase-H, sliding clamps, and other factors (31). Considerably less is known for archaeal replication, where mostly B type polymerases, similar to eukaryotic replication enzymes pol α and δ, have been identified (6, 22, 23, 32–34). The criteria appreciate it surprising, because such crucial biotechnological applications as cloning and PCR require the thermostability and fidelity typical of archaeal polymerases (35). Thus, in addition to satisfying basic research interests, structural information could aided, for example, the engineering of variant enzymes with tailored nucleotide incorporation rules or the design of antiviral and anticancer polymerase inhibitors. For these reasons, we have determined the structure of a DNA polymerase from *Thermococcus gorgonarius* (Tgo), an extremely thermophilic archaeal archaebacteria isolated from a geothermal vent in New Zealand (36). The enzyme possesses pol and $3' \rightarrow 5'$ exonuclease activity, which together ensure thermostable replication with high fidelity (error rate: $3.5 \times 2.3 \times 10^{-6}$; see ref. 36). The 2.3 Å resolution shows a topological similarity to gp43 and gives insight in the structural biology of archaeal DNA polymerases, including the identification of special conclusions for thermophilic adaptation.

MATERIALS AND METHODS

Materials. All materials were of the highest grade commercially available.

Bacterial Strains. *Escherichia coli* BL21(DE3) containing pUBS520 was used as described (36). *E. coli* JM109 (DE3) (Promega) was a generous gift of Matthias Bochtler (Max-Planck-Institut).

Expression Vectors. PET21d was obtained from Novagen Molecular Biochemicals.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: Tgo, *Thermococcus gorgonarius*; DNA pol, DNA polymerase; Gaussian, The atomic coordinates have been deposited in the Protein Data Bank, Brookhaven National Laboratory (Upton, NY) (ref. 59) (PDB ID code 1TGO).

‡Present address: Gene Center, Ludwig-Maximilians-Universität, Feodor-Lynen-Strasse 25, D-81377 München, Germany.

§Present address: Department of Biochemistry, La Jolla, CA 92161.

||To whom reprint requests should be addressed. e-mail: bernhard@roche.com.

Table 1. Data collection and nonreference replacement

Derivative	Resolution limit, Å	Total observations	Unique observations	Completeness, %	R_{sym}, %	R_{iso}, %	Phasing power
Native	3.0	136,915	21,379	91.9	7.9		
LP	3.2	75,964	20,119	99.5	11.3	18.1	0.71
PTP	3.2	69,836	17,955	96.2	12.0	27.5	1.34
PTP	4.0	33,870	6,910	68.9	14.3	12.8	1.17
PTC	3.4	90,336	14,105	93.8	10.5	17.1	1.16
PTA	3.5	107,827	15,489	98.6	7.7	18.5	0.87
PTT	2.3	328,695	26,585	88.7	6.1	36.2	0.83
PTH	3.5	39,629	11,457	92.4	9.6	38.7	1.24
PD	3.5	78,186	13,206	97.9	11.7	17.5	0.43
PMPT	3.4	94,532	14,155	99.6	8.9	19.6	1.04
OS	2.8	488,168	24,036	91.2	5.2	28.4	1.81

Outermost layers of each (3.0–3.0 Å). 0.79

Heavy-atom derivatives were prepared by soaking the crystals in low salt buffers containing the heavy atoms in solutions 0.5–0.8 mM except for 2 h, PTA, 3 mM K$_2$PtCl$_4$ for 1 h; PTT, 3 mM K$_2$PtCl$_4$ for 2 d; PTT, saturated room-temperature solution of TlNO$_3$ for 1 d; PTH, 2 mM K$_2$PtCl$_4$ for 1 d; PTS, saturated room-temperature solution of $(NH_4)_2SO_4$ for 3 d; PMPT, PD + PTS; OS, saturated K$_2$PtCl$_4$ for 1 d. MAD, PTP, PTA, PTS, PTC, PTH and PMPT were collected with a MAR imaging plate and MAR CCD (Mar Research, Hamburg). All other data sets were collected with a Mar-design cryogenic device at beamline X06SA of the SLS (PSI, Villigen). All other data sets were collected with a Mar-design cryogenic device at beamline BW6 at DESY, Hamburg.

Cloning, Expression, and Protein Purification. The gene for the MSH ABC DNA-dependent pol (Tgo pol) was cloned from *Thermococcus gorgonarius* (see Klimczak and Kulbachinskaya, ref. 19) and expressed in *E. coli* LE392pUBS520 pRos2Tgo (Escherichia coli-based system with transcription activator, UBS520) as described (20). Tgo pol was purified essentially as described (20) with the substitution of the TSK butyl Toyopearl column by Blue-Trisacryl M (Tm-tri) and with an additional concentration step on Poros 50 HQ anion exchange medium (Perseptive Biosystems, Framingham, MA). Active fractions were combined, concentrated to 20 mg/mL, and transferred to 20 mM sodium phosphate pH 8.2/10 mM 2-mercaptoethanol/500 mM NaCl.

The gene for a subunit of thermosome-containing subunit of Tgo pol (the Tgo pol) was expressed in *E. coli* BL21 (DE3) (Invitrogen) using established protocols (17). The Tgo pol was purified by using the acid cycle protocol.

Crystallization. Crystals of purified Tgo pol (or Sa-Tgo pol) were grown by using sitting drop vapor-diffusion technique in 4°C with high-salt conditions [(a) as protein-precipitant—100 mM Tris, pH 8.2/200 mM sodium sulfate; and cultivated to 3.0 Å (freezing) and to 2.8 Å (room-temp.) with a Rigaku (Danvers, MA) R-axis IIc/RU-200 rotating anode source, or with a MAR imaging plate on a MAR CCD (charge-coupled device) at beamline BW6 at DESY, Hamburg. The data were processed with DENZO (Molecular Applications, Inc., Corte Madera, CA) or MOSFLM (MAR imaging plate) ref. 30), scaled with SCALA (ref. 30) or SCALEPACK (30), and reduced with TRUNCATE (30).

Structure Determination. The structure was solved by multiple isomorphous replacement and anomalous scattering (MIRAS) by using data from crystals transformed to low salt conditions (Table 1). Crystallographic calculations were done with programs from the CCP4 suite (40). Heavy-atom positions of major sites were located in difference Patterson maps and were refined and improved (ref. 30; see also ref. 41). Phases were calculated with the program MLPHARE (30), and solvent flattening and histogram matching (30) were followed by automatic model building with ARP/wARP (42). Model building was performed with TURBO-FRODO (43) on Silicon Graphics workstations (Mountain View, CA) and with O (44) on Linux PCs. Refinement was performed with X-PLOR (45) and CNS (46). Model building and refinement were performed in cycles. Stereochemistry was checked with PROCHECK (47). Figures were prepared with MOLSCRIPT (48), RASTER3D (49), and GRASP (50).

Sequence Determination. The sequence was determined by multiple isomorphous replacement and anomalous scattering (MIRAS) by using data from crystals transformed to low salt conditions (Table 1). Crystallographic calculations were done with programs from the CCP4 suite (40). Heavy-atom positions of major sites were located in difference Patterson maps and were refined and improved (ref. 30; see also ref. 41). Phases were calculated with the program MLPHARE (30), and solvent flattening and histogram matching (30) were followed by automatic model building with ARP/wARP (42). Model building was performed with TURBO-FRODO (43) on Silicon Graphics workstations (Mountain View, CA) and with O (44) on Linux PCs. Refinement was performed with X-PLOR (45) and CNS (46). Model building and refinement were performed in cycles. Stereochemistry was checked with PROCHECK (47). Figures were prepared with MOLSCRIPT (48), RASTER3D (49), and GRASP (50).

Cloning, Expression, and Protein Purification. The gene for the MSH ABC DNA-dependent pol (Tgo pol) was cloned from *Thermococcus gorgonarius* (see Klimczak and Kulbachinskaya, ref. 19) and expressed in *E. coli* LE392pUBS520 pRos2Tgo (Escherichia coli-based system with transcription activator, UBS520) as described (20). Tgo pol was purified essentially as described (20) with the substitution of the TSK butyl Toyopearl column by Blue-Trisacryl M (Tm-tri) and with an additional concentration step on Poros 50 HQ anion exchange medium (Perseptive Biosystems, Framingham, MA). Active fractions were combined, concentrated to 20 mg/mL, and transferred to 20 mM sodium phosphate pH 8.2/10 mM 2-mercaptoethanol/500 mM NaCl.

The gene for a subunit of thermosome-containing subunit of Tgo pol (the Tgo pol) was expressed in *E. coli* BL21 (DE3) (Invitrogen) using established protocols (17). The Tgo pol was purified by using the acid cycle protocol.

Crystallization. Crystals of purified Tgo pol (or Sa-Tgo pol) were grown by sitting drop vapor-diffusion technique in 4°C with high-salt conditions [(a) as protein-precipitant—100 mM Tris, pH 8.2/200 mM sodium sulfate; and cultivated to 3.0 Å (freezing) and to 2.8 Å (room-temp.) with a Rigaku (Danvers, MA) R-axis IIc/RU-200 rotating anode source, or with a MAR imaging plate on a MAR CCD (charge-coupled device) at beamline BW6 at DESY, Hamburg. The data were processed with DENZO (Molecular Applications, Inc., Corte Madera, CA) or MOSFLM (MAR imaging plate) ref. 30), scaled with SCALA (ref. 30) or SCALEPACK (30), and reduced with TRUNCATE (30).

RESULTS AND DISCUSSION

Structure of Tgo pol. Tgo pol is a long chain of 775 amino acids (Fig. 2). The N-terminal domain (residues 1–280, the 3′ → 5′ exonuclease domain (131–328), the palm (369–499) and thumb (500–582), fingers (600–695), and thumb (780–775) domains of the polymerase unit, and a helical intermediate insertion (327–368)

Table 2. Crystallographic refinement, Mpkmol.dom

Item	Value
Space group	$P2_12_12_1$
Cell dimensions, Å	a = 58.1, b = 105.2, c = 154.2
Observations, 25–2.5 Å	
Total	452,446
Unique	39,421
Completeness, %	
Total	93.1
Last shell	86.6
R_{sym}, %	
Total	7.3
Last shell	20.2
R factor (R_{free}), %	20.9 (25.1)
rms deviation in bond lengths, Å	0.008
rms deviation in bond angles, °	1.5
No. of nonhydrogen atoms	
Protein	6,759
Water	320

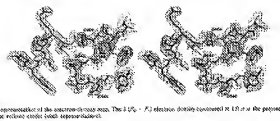


FIG. 1. Stereoscommonstrator of the electron-density map. The 3.0Å – F_o electron density (contoured at 1.0 σ) at the polymerase active site is well-defined by the refined atomic model superimposed.

between the exonuclease and palm domains. The polymerase cleft lies on the DNA-binding crevice, reminiscent of a right hand, which is the overriding characteristic of pols. Opal 3 from bacteriophage RB69 also shows this overall domain topology (15).

These clefts extend radially from the polymerase active site at the center of the ring; two of them, in opposite directions, forming a major cleft across the molecule, and one perpendicular to those. Based on sequence homology to pol A family enzymes, the two opposite clefts probably bind duplex DNA (cleft D, according to ref. 15) and single-stranded template DNA (cleft T), respectively. The perpendicular (trifling) cleft binds the polymerase active site and the exonuclease active site and breaks the purine strand in editing mode (11).

The exonuclease domain is structurally equivalent to the 3' → 5' exonuclease domain of pol A (Klenow fragment of pol I, Escherichia coli) (16). Like gp43, however, it is located the opposite side of the polymerase cleft by approximately 40 Å. The exonuclease domain is connected to the palm domain by a pair of anti-parallel α-helices (residues 187–208 and 211–232). The exonuclease domain contacts the palm domain at the N-terminal and palm domains and the 42 residues interdomain helix, on the other side. This latter segment is located at the base of cleft T, which is additionally bounded by the exonuclease, 70 Å distant, and palm subdomains.

The topology of the palm domain is conserved among polymerases (Klenow (5), with two long helices (O and R)

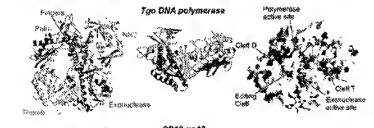


FIG. 2. Structure of Tgo pol and comparison with gp43 from bacteriophage RB69. Half-tone ribbon representation of Tgo DNA polymerase (figure at two orthogonal views with labelled secondary structure elements). The molecule is composed of five domains. N-terminal domain (yellow), 3' → 5' exonuclease (red), palm (light and dark magenta), fingers (dark and light green), and thumb (green), which are arranged to form a ring. An interdomain α-helix (green) between the exonuclease domain and the palm is strange. The conserved carboxylates in the palm site and the two ajacular cations are shown as red and yellow balls, respectively. Tgo pol has the same overall architecture and domain topology than gp43 of RB69 (lower). The fingers are colored in the figure, as gp43 is grey. Bright Compressed molecular surfaces of Tgo pol (upper) and RB69 gp43 (lower). Red and blue indicate negative and positive electrostatic surface potentials, respectively. In contrast to gp43, Tgo pol has a strongly enhanced positive potential at the putative DNA binding cleft.

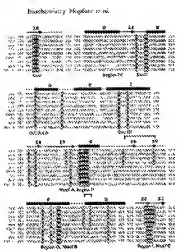


Fig. 3. Sequence alignment of B family DNA polymerases. The alignment was built in shaded boxes (of 3) to highlight specific residues from the most conserved segments. The secondary structure of Tgo pol is indicated on top of the alignment with helices (black), strands (yellow) and loops (thin) colored according to domains with the same color code as Fig. 1B. Strictly conserved residues of type B polymerases are red, and underlined conserved residues yellow. Universally conserved residues of archaeal type B enzymes are underlined in the same color code. Dots indicate gaps. Abbreviations: Tgo, *Thermococcus gorgonarius* pol; Pfu, *Pyrococcus furiosus* pol; Tpe, *Thermococcus* sp. pol. 9°N; Tli, *Thermococcus litoralis* pol; Mth, *Methanobacterium* sp. pol; Mfu, *Methanococcus fulgidus* pol; Hbu, *Halobacterium* sp. NRC-1 pol; Pab, *Pyrococcus abyssi* pol; Ec, coli pol II; Sce, yeast pol II.

interlock against the five-stranded antiparallel β-sheet that provides the three conserved aspartate residues involved in nucleotidyl transfer. The fingers emerge from the palm domain as an α-helix-rich insertion. Its 50 residues are folded into two segments that embed another set of opposite-handed α-helices. Helix F contains the conserved Ka/Re/Y/Q-x(2)-motif of B type polymerases and is inserted in the O helix of A type enzymes (see below). The ~26-residue insertion between helix O and F in B/RB69 and T4 pol's is missing in Tgo pol, whose helix bridges and connects these two much shorter than their equivalents in pol's. The shorter fingers of Tgo pol presumably reflect the typical structure of the archaeo/eukaryote B type fingers (not to mix A- and B- cell pol II). The thumb domain topology, similar to that of pol's, is unrelated to other polymerase types. However, like the thumb of A type enzymes, a bundle of α-helices set in two grooves from the screw into β-sheet fixture in the screw fits the thumb contains a ~75-residue α-helix-rich (α65–72%), which fixes the unstructured domains and contributes to the tubing channel, explaining why mutations in the noncatalytic domain of B-type polymerases affect the polymerase activity and vice versa [44, 45].

Weakly defined density across the tunnel of the thumb domain was modeled as the C-terminal 6 residues with a polyalanine chain. The C terminus must then run proximate from the same molecule as in the RB69 polymerase [21]. Because the C terminus of the T4 pol was involved in sliding-clamp binding [45], it is likely, however, that these residues become ordered on the same binding to the factor.

Sequence Alignment of Archaeal DNA Polymerases. The structure of Tgo pol allows the generation of a structure based sequence alignment of the archaeal subfamily of type B DNA polymerases, the location of conserved and unique residues, and the comparison with other type B DNA polymerases (Fig. 3).

Polymerase Active Site. The polymerase active site is formed by the central β-sheet extends 16, 17, 22, 11, and 22 into helix M of the palm domain and helix F located in the fingers and is highly conserved among B family polymerases (Fig. 4). Three carboxylates required for nucleotidyl transfer in B family polymerases, two of which coordinate two metal ions [30] are exceptionally conserved among A family enzymes, B family enzymes, and reverse transcriptase [31]. Superposition of Tgo pol and T7 replication complex [19] places the dNTP near the proposed nucleotide-binding site in helix F, Thr 548/Ala/Ser/Thr/pol motif (Fig. 3) may suggest interaction of the carbohydrate with residue and the phosphate tail of the bound dNTP (Fig. 4). Protonation of the catalytic conserved Lys-487 allows it to ionize the Lys-522-phosphate tail interaction in T7. Tyr-404 (Ile/Phe/Cys/Ala/Gly) and Tyr-409 (Ile/Val/Phe/Ile) form the bottom of the nucleotide binding site.

The active site of B family polymerases contains a DTDS motif, which, however is DTDG in the enclosed subdomain. In the pol, the substrate symmetric Tyr-402 from the adjacent strand provides an electron withdrawal group, or a proximal oxygenation for metal coordination or binding of the 3′ end of the primer. The orientation of Tyr-402 is stabilized by the aromatic cluster that also encloses Pro-361 and Tyr-508. Archaeal *Methanococcus vulnae* and *Thermococcus* sp. 9°N (see Fig. 3) have Tyr or position 508—Phe in Tgo—earlier than at 402, but might also satisfy an aromatic group. The displacement of a functional element from serine in DTDS to Tyr-402 or Tyr 508 might stabilize its conformation as an adaptation for exo-nucleolysis.

The conserved cluster of amino acids DSTR, EMRE form an unexpected metal-binding site for Mn^{2+} and Zn^{2+} (Fig. 4). Its proximity to Asp-487 and to the conserved location of the dNTP γ-phosphate suggests a supporting role in nucleotide binding and/or catalysis.

3′ → 5′ Exonuclease Active Site. Tgo pol is characterized by a strong 3′ → 5′ exonuclease activity, unlike eukaryotic B type polymerases (unpublished results). The exonuclease active site is formed at the interface between the catalytic loops domain and the tip of the thumb (Fig. 5). All residues important for catalysis are located in the noncatalytic domain, which, at least for T4 [45], retains activity when dissociated from the polymerase [47]. However, the thumb domain, with, for example, Ribo/gp45 Phe 120–here intercalation, partially interlocks the hydrolytic geometry of single strand DNA [21, 45].

The exonuclease structures of Tgo and gp43 DNA polymerases are similar at the editing site but differ considerably at the exonuclease–thumb interface. Mutant 19 restores the vestal locking D156E motif and notably superimposes with the exonuclease strand from gp43, allowing mobility of a single-strand DNA segment into the exonuclease site from the backbone/gp43 Phe 120–base intercalation, partially interlocks the hydrolytic geometry of single strand DNA [21, 45].

The exonuclease structures of Tgo and gp43 DNA polymerases are similar at the editing site but differ considerably at the exonuclease–thumb interface. Mutant 19 restores the vestal locking D156E motif and notably superimposes with the exonuclease strand from gp43, allowing mobility of a single-strand DNA segment into the exonuclease site from the backbone/gp43 Phe 120–base intercalation, partially interlocks the hydrolytic geometry of single strand DNA [21, 45].

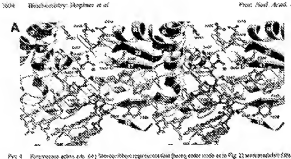


FIG. 4. Polymerase active site. (A) Stereochemical representation (same color code as in Fig. 2) with modified DNA. Amino acid residues are shown in ball-and-stick representation, with carbon (green), nitrogen (blue), oxygen (red) atoms. The DNA template (light brown), primer (light brown), and incoming nucleotide (gray) are from the crystal structure of T7 replication complex (47) by superimposing DNA. Side and adjacent residues with corresponding residue in T7 gp5 (D903 and E907) are labeled. Phosphorus atoms are yellow. The two metals of the T7 replication complex are shown as magenta spheres. (B) Experimentally observed model showing the Fpg Mn^{2+} (K_d = 4) formally positioned at Sn2 and Zn^{2+} at the "low site" crystal sites. The carboxylates E57Q and E260 are conserved in type II polymerases (Fig. 3).

947). Moreover, the sliding clamp is demarcated by a depression at the lip of the protein toward the intracellular domain to grab the tether strand (Fig. 5B). This shelf is constructed with a large charge at the loop between strands 3B and 3C in RB69 gp45 (and likewise in T4 gp45), the loop formed by residues 77–87 have not been seen (see Phe-123, which immediately follows the first two bases). In Fpg and, the loop is turned backward, away from the tether, and the α5C helix (the equivalent of gp45's Thr-123) attaches to Phe 234 30 Å away from intercalation site. This shift allows the tip of the thumb to move into the sliding channel and to block the translocation site.

Are There Different Conformations in Polymerase and Editing Modes? If a closed conformation of the nonextension duplex product might strand binding, an open conformation is necessary for editing. The observed distinct conformations may represent the extreme in "polymerase" mode. Preliminary analysis of the crystal structure of Fpg not in the low-salt conditions indicates a structural change at the interface of subdomains and thumb, possibly reflecting a transition between open and closed forms. The closed conformation observed here may, however, be a nonphysiological artifact of the high ionic strength used for crystallization. Control structures of the enzyme in both polymerase and editing modes are required.

Adaptation to High Temperatures. Fpg in a sulfur-metabolizing, extremely thermophilic crenarchaeon, with a growth range between 75°C and 98°C. For extensive replication at this temperature range, the polymerase must not only be stable, but must also adequately bind substrate DNA. A comparison with gp45 from the mesophilic bacteriophage RB69 indicates several such adaptations to high temperatures. Several loops are shortened. Fpg gp45 has 6 gp45 (Fig. 2), and there is an increase in hydrogen bonding at crystal contacts. Fpg gp45 secondary structure includes 40% helix, 22% B-strands, and 19% turns (calculated according to ref. 48), whereas gp45 has 42% helix, 17% B-strands, and 18% turns.

Although very strong cytoplasmic or nuclear proteins, two disulfide bridges (open in present reaction pairs 428–442 and

FIG. 5. 2′- or 3′-monophosphate active site. (A) Stereochemical representation with modified DNA using the color code of Fig. 4A (ball and stick) and Fig. 2 (ribbons). Amino acid residues are shown in ball-and-stick representation. The single-stranded DNA backbone follows from the crystal structure of RB69 single-stranded DNA complex (49). The arrangement of the DNA has been simulated by superimposing the E683A1 motif in Tpo pol with corresponding motif of RB69 gp43 (2PHL168). Bound 2′ and 3′ nucleotide products from the thymine (green) apparently in collision with the modified DNA. (B) Comparison of the monomeric thumb interface between Tpo pol (same color as Fig. 1B) and RB69 gp43 (gray). In Tpo pol, the tip of the editing site motif is here oriented consistent with the equivalent loop of gp43 (yellow), allowing the tip of the thumb to move toward A (N$_2$ close to the exonuclease. This conformation is incompatible with formation of an editing complex (like gp43's) of specific chosen sequence single-strand tether.

106–109, although reduced, are present for recombinant (Fig. 2). Mutant enhancement and alkaline density dispersion with specific centrifugation for the double heterozygous Rad mutant (data not shown) demonstrated 363-fold decrease. 2.8 Å and 3.0 Å. This is consistent with our E. coli expression and further indicates that actual perturbation by nonreducing conditions. These cysteines are located in the globe domain and are conserved among B type enzymes from hyperthermophilic archaeon, but not among mesophilic homologs (Fig. 3). The Cys-426–Cys-441 bridge stabilizes the compact fold of the loop segment between helix N in the globe domain and helix O in the flagpole and presumably aids the selective interactions of these helices. In addition, the loop segment packs against helix Q in the pore domain. Helix Q, the core of the pore domain, is the first α-helix of the first half-subunit by the second dimer-like bridge between Cys-306 and Cys-309.

A much enhanced complementarity positive potential for all three DNA-binding clefts of Tpo pol is observed relative to pol1 (Fig. 2). Thus, in addition to hydrogen bonding and specific DNA–protein interactions, binding to Tpo pol has an additional strong, stabilizing electrostatic component.

An attribute in the position of salt bridges is often associated with thermostability. Although Tpo pol has a greater total number of charged residues (202) than pol1 (242), both molecules have 54 salt bridges within a 5-Å distance. However, at the 5–7 Å range of charge distance, Tpo pol has 77 ion pairs compared with 42 for pol1. This large increase results in a more highly charged surface of Tpo pol, accompanied by a more balanced charge distribution, compared with pol1 where charges are often located in patches (Fig. 2).

We thank Dr. Uwe Baxa and Mandy Augustin for introducing crystallization, Dr. Stefan Weitz for help with expression of heterologous proteins, and Dr. Hans Eberhard and the GBF fermentation group for help with data collection.

1. Wang, J., & Lehman, I. R. (1989) *Annu. Rev. Biochem.* **67**, 721–751.
2. Mizuta, K., & Ishikawa, H. (1989) *Nucleic Acids Res.* **17**, 9277.
3. Wang, T. S.-F. (1991) *Annu. Rev. Biochem.* **60**, 513–552.
4. Sakaguchi, A. & Sakaki, Y. A. (1992) *DNA Replication* (Freeman, New York).
5. Ito, J., & Braithwaite, D. K. (1991) *Nucleic Acids Res.* **19**, 4045–4057.
6. Braithwaite, D. K., & Ito, J. (1993) *J. Mol. Biol.* **231**, 241–267.
7. Braithwaite, D. K., & Ito, J. (1993) *Nucleic Acids Res.* **21**, 787–802.
8. Argos, P. (1988) *Nucleic Acids Res.* **16**, 9909–9916.
9. Poch, O., Sauvaget, I., Delarue, M., & Tordo, N. (1989) *EMBO J.* **8**, 3867–3874.
10. Ito, J., & Braithwaite, D. K. (1991) *Nucleic Acids Res.* **19**, 4759–4766.
11. Bernad, A., Blanco, L., Lazaro, J. M., Martin, G., & Salas, M. (1989) *Cell* **59**, 219–228.
12. Blanco, L., Bernad, A., Lazaro, J. M., Martin, G., Garmendia, C., & Salas, M. (1991) *J. Biol. Chem.* **266**, 8136–8141.
13. Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., & Angerer, B. (1999) *J. Mol. Biol.* **289**, 1135–1150.
14. Kong, H., Kucera, R. B., & Jack, W. E. (1992) *J. Biol. Chem.* **267**, 1350–1356.
15. Braithwaite, D. K., & Ito, J. (1993) *Nucleic Acids Res.* **21**, 5856–5864.
16. Rodriguez, A. C., Park, H.-W., Mao, C., & Beese, L. S. (1991) *J. Mol. Biol.* **299**, 447–462.
17. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., & Kraut, J. (1994) *Science* **264**, 1891–1903.
18. Kiefer, J. R., Mao, C., Braman, J. C., & Beese, L. S. (1998) *Nature (London)* **391**, 304–307.
19. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., & Ellenberger, T. (1998) *Nature (London)* **391**, 251–258.
20. Johnson, A. & O'Donnell, M. (1998) *Annu. Rev. Biochem.* **64**, 133–168.
21. Kong, X.-P., Onrust, R., O'Donnell, M., & Kuriyan, J. (1992) *Cell* **69**, 425–437.
22. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
23. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
24. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
25. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
26. Fujikawa, K., Kobayashi, H., Tomizawa, J., & McMacken, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7734–7738.
27. Kornberg, A. & Baker, T. A. (1992) *DNA Replication* (Freeman, New York).
28. Jager, J., Sandman, K. E., Pereira, S. L., Lurz, R., & Reeve, J. N. (1998) *Biochemistry* **37**, 2811–2818.
29. Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H., & Steitz, T. A. (1989) *Cell* **89**, 1087–1099.
30. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
31. Kong, X.-P., Onrust, R., O'Donnell, M., & Kuriyan, J. (1992) *Cell* **69**, 425–437.
32. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
33. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
34. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
35. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
36. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
37. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
38. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
39. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
40. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
41. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
42. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
43. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
44. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
45. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
46. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
47. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
48. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
49. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
50. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
51. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
52. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
53. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
54. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
55. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
56. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
57. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
58. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
59. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
60. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
61. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
62. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
63. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
64. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
65. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
66. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
67. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
68. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
69. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
70. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
71. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
72. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
73. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
74. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
75. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
76. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
77. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
78. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
79. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
80. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
81. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
82. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
83. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
84. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
85. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
86. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
87. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
88. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
89. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
90. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
91. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
92. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
93. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
94. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
95. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
96. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
97. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.
98. Kelman, Z. & Hurwitz, J. (1998) *Trends Biochem. Sci.* **23**, 236–238.
99. Bruck, I., Perron, G. G., Dixon, N. E., & Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 20873–20881.
100. Delagoutte, E. & von Hippel, P. H. (1998) *Biochemistry* **37**, 1402–1414.

Appendix 1

We have purified and characterized the Family BDNA polymerase from the archaeon Methanococcus maripaludis, cloned from ATCC 43000. This polymerase has a 41% sequence identity and 63% sequence similarity with Vent DNA Polymerase when analyzed using NCBI Blast 2 and the default parameters.

We performed the titration assay described in Example 1 of the patent application, using the Mma, Vent (exo-), and 9°N (exo+) DNA Polymerases. Experimental details and data are given in the attached figure.

For each of the three polymerases, a comparison of lanes using dideoxyCTP (ddCTP) with those using equivalent concentrations of acycloCTP (acyCTP) reveals shorter products in lanes utilizing acyCTP. These shorter products result from more efficient insertion of the acyCTP terminator compared to incorporation of the ddCTP terminator. Thus, all three polymerases incorporated acyCTP more efficiently than ddCTP.

Figure Legend

The ability of acyNTPs and ddNTPs to act as chain terminators was tested using a titration assay of the type described in Example 1. Incorporation of ddCTP was compared to that of acyCTP, respectively, using Methanococcus maripaludis DNA polymerase, 9°N (exo+) DNA polymerase and Vent® (exo-) DNA polymerases.

Incorporation of ddCTP and acyCTP was assayed by mixing 8 µl of reaction cocktail (0.025 µM 5' [FAM] end-labeled #1224-primed M13mp18, 62.5 mM NaCl, 12.5 mM Tris-HCl (pH 7.9 at 25°C), 12.5 mM $MgCl_2$, 1.25 mM

dithiothreitol, *Methanococcus maripaludis* DNA polymerase or 0.125 U/µl 9°N (exo+) DNA polymerase or 0.125 U/µl Vent® (exo-) DNA polymerase) with 2 µl of 5X nucleotide analog/nucleotide solution to yield the final ratios of analog:dNTP indicated in the figures. After incubating at 72°C for 20 minutes, the reactions were halted by the addition of 10 µl formamide. Samples were then heated at 72°C for 3 minutes and a 1 µl aliquot was loaded on a 4% polyacrylamide urea gel and detected by an ABI377 automated DNA sequencer.

ddCTP v. acyCTP incorporation by archaeal DNAPs